Abstract.

Most scientific models are idealised and simplified representations of their targets at least to some extent, but some models are very simple and or highly idealised. Yet these simple models are often claimed to help scientists to understand the world of their scientific theories. How simple models might do this is the focus of this thesis. I will first give an overview of different types of simple models. Next, I will discuss on the notion of understanding as is appropriate to this project. Then I will present and consider several different ways simple models are claimed to help scientists understand, including by providing explanations of various sorts and by elucidating concepts.

TITLE:

Simple models and understanding in science

By Uzma Malik

MA by research

Department of Philosophy,

Durham University, 2020.

Table of contents:

List of tables:

List of illustrations:

1. Introduction

What are scientific models? Scientific models can be [at least] physical material objects, fictions, descriptions, set theoretical structures or mathematical equations (Frigg and Hartmann 2012). It is often assumed that models *represen*t i.e., stand in some representational relation to a target system. Moreover, scientific practice has revealed the pervasive use of *idealised* and often quite simple models in the investigation of natural and social phenomena, even in cases where such phenomena are thought to be quite complex.

Highly idealised and simple models are often claimed to promote *understanding in science*. For instance, philosopher Angela Potochnik says that such models "regularly have direct epistemic value even at the cost of accuracy, in so far as they promote understanding" (Potochnik 2020 p.934). This is somewhat puzzling: how can these very simple and/or idealised models, which fail to accurately represent their targets in various ways, nevertheless facilitate understanding ? After all, understanding, when it is discussed in philosophy of science at all, is often thought to be related to explanation; and explanation is often related to truth, as in Hempel's influential Deductive-Nomological model.

In fact, on Hempel's view, understanding was a subject of inquiry for the psychological sciences or the history of science and was not seen as an epistemic aim of science (De Regt 2013). For a long while, it was a topic given little attention by philosophers. Recently, however, there has been a resurgence of interest in the topic of understanding. Whilst there are now many different views of understanding, many of them do leave room for simple and idealised models to have an important role in advancing understanding, as I will discuss in this thesis.

Exactly how simple and idealised models facilitate scientific understanding is a relatively new area in the philosophy of science. Broadly speaking, this thesis provides a survey of extant views regarding how they can do this. The aim of this thesis is to organise, interpret and synthesise the relevant literature -- to arrive at a more comprehensive perspective on how simple models of various types can facilitate understanding.

In the second chapter, the focus is on simple models. I characterise scientific models and discuss simple scientific models as a subclass of scientific models. I then explore several types of simple scientific models. These are: highly idealised and very simple *toy models*; simple models that perform an *epistemic surrogate function; minimal models* that aim to not let the details get in the way; and *target-less models* that do not represent systems of interest but relate to them by analogy or similarity.

In the third chapter, I explore key features of contemporary views of the nature of understanding and clarify what I will assume about understanding in the remainder of the thesis. My general assumptions after surveying the literature will be something that facilitates understanding need not be entirely factive (i.e., accurate and true); it will usually facilitate understanding in part via explanation; obtaining understanding will involve some kind of grasping of content; and the understanding gained maybe any of several "kinds": objectual understanding, understanding "with", understanding "that" or understanding "how". I will assume that a model facilitates understanding if it facilitates any one of these kinds of understanding.

The following chapters, chapters 4-6 inclusive, consider different answers to the question of how simple models facilitate understanding for the various categories of simple models.

In the fourth chapter, I consider the view that simple models can facilitate understanding by providing how-actually and how-possibly explanations that researchers can grasp. Different philosophers make the how-possibly and how-actually distinction in different ways, though Reutlinger et al. (2018) provide the overarching conception, positing two distinct types of

explanations: the first one is a how-possibly explanation (which can also be characterised as a partial how-actually explanation) and the other is a how-actually explanation. Both Bokulich (2014) and Grune-Yanoff (2016) have distinct views about the ways in which these how-possibly and how-actually explanations interact with each other, which can be viewed as in effect extending or elaborating the Reutlinger et al. (2018) conception.

In the fifth chapter I discuss Batterman's (Batterman 2002, 2009, 2018, Batterman and Rice 2014) view of how minimal models facilitate explanation and thereby understanding. I articulate a general version of Batterman's method of minimal model explanation using paradigmatic renormalisation group methods. Then I present a more detailed example from physics as well as an example from biology to show that Batterman's minimal model explanation can be applicable to other fields in science. Lastly, I link minimal model explanation to understanding in greater detail than Batterman, discussing unification and contrasting this with the mechanistic variety. I characterise Batterman's minimal model explanation as subscribing to a kind of "local" unificationist understanding.

In the sixth chapter, I turn to target-less simple models and a different variety of understanding known as objectual understanding. Following Luczak (2017), I show one way in which the Kac ring can contribute to objectual understanding of statistical mechanics. It does this by elucidating the recurrence and reversibility objections in statistical mechanics. I show that, in doing this, it fulfils four conditions for objectual understanding as outlined roughly by Baumberger (2019) and Baumberger and Brun (2016). So, I suggest the Kac ring model helps us to get a better, objectual understanding of theory in statistical mechanics.

The conclusion of this thesis states that there is much to be done in terms of further work—I outline some areas for further work. I also articulate what I think are the main contributions of each chapter to the thesis: how simple models facilitate understanding.

2. What are simple models?

2.1. Introduction

This chapter discusses the nature of scientific models and, in particular, simple models. In section 2.2, I articulate common characteristics of models in science. In section 2.3, I introduce idealisation. In section 2.4 I introduce simple models and contrast them with complex models. In section 2.5, I discuss four types of simple models: toy models 2.5.1, epistemic surrogates 2.5.2, minimal models 2.5.3, target-less simple models 2.5.4. The types are not mutually exclusive and exhaustive but each highlight important features of at least some simple models.

2.2. Scientific models

In this section, the aim is to offer a view of what scientific models are. In general, a scientific model is a simplified and idealised representation of a target system; it represents one or more aspects of the world that we are interested in. Models are common in all scientific domains, thought to be useful for prediction, explanation, teaching, heuristic, and other purposes.

From an ontological perspective, models can be, at least: physical or material objects, fictions, descriptions, set theoretical structures, or mathematical equations (Frigg and Hartmann 2012). An example of a physical model is the original Schelling model of racial segregation which used a checkerboard and some coins. Coins represented real people or households and their configuration represented neighbourhoods, with the checkerboard squares representing spatial locations in a city.

Another well-known example of a physical model is Watson and Crick's model of DNA. An example of a fictional model is a frictionless pendulum, which is meant to represent an object in the scientist's mind, because there is no such thing as a frictionless pendulum in the real world. An example of a model as a set of equations is the Mundell-Fleming model of the open economy, whereby the equations that represent the open economy are syntactic items (paraphrased from Frigg and Hartmann 2012). I am primarily concerned in this thesis with mathematical models that have targets i.e., are meant to represent some real or imagined system.

2.3. Idealisation

It can be said that all scientific models are simplified and idealised, at least to some extent. A model can be simplified in the sense that it represents a reduced number of causal or explanatory factors at work within the target system or represents them with reduced detail. For example, not all factors that affect the Earth's climate are represented in climate models. Simplicity comes in degrees ranging from slightly simplified to extremely simple. Similarly, if a model is idealised then it contains idealisations. Idealisations assert something literally false of the target system, so at first glimpse, these models do not accurately represent their target system (Reutlinger et al. 2018). For example, an economic model might assume that actors have perfect information, which is false, as no real actor has perfect information. How idealised a model is can also come in degrees and an idealised model can have different types of idealisations.

Similarly, there are (at least) two types of idealisation that are often distinguished in the literature (Weisberg 2013, Frigg and Hartmann 2012): Aristotelian and Galilean.

Aristotelian idealisation strips away some of the features of the target model. It strips away some of those causal factors that affect the target and focuses on others. For instance, a bob's mass in an idealised pendulum is treated as a point mass, and so its shape, volume, colour, et cetera, are stripped away to focus on the characteristics of the model that one wants to investigate. This is a general characterisation of Aristotelian idealisation which is given further nuance and descriptions by different authors (e.g., Cartwright 1983 and 1999). Sometimes Aristotelian idealisation is also called minimal idealisation (Weisberg 2013 pp.100-103). Such minimalist representations are often associated with the aim of understanding as discussed further in later chapters.

Aristotelian idealisation may be contrasted with Galilean idealisation. Galilean idealisation may be characterised as a deliberate distortion of the target system often to promote mathematical or computational tractability (Frigg and Hartmann 2012); it is hoped that idealisations will eventually be removed from the model, replaced with less distorted representations. An example of Galilean idealisation can be found in computational chemistry; the aim is to calculate molecular properties by computing approximate wave functions for the molecules of interest. More complex or chemically interesting systems are handled with the increase of computational power; today It is possible to compute extremely accurate but still approximate wave functions for moderate sized molecules. Eventually the aim is to calculate the exact solution to the Schrödinger equation; that is the limit to which all the reduction in approximation aim towards (paraphrased from Weisberg 2007 p.642).

It is often assumed that idealisation relates to a particular aspect of a model. However, there is another kind of view on which some idealised models in science are holistically, that is, pervasively distorted representations of the target system (Rice 2019). These idealised models are said to be greater than the sum of the accurate and inaccurate (idealised) parts, resulting from a complex interaction of various modelling assumptions. Such holistic distortion is justified because it allows for the application of mathematical modelling techniques that provide epistemic access to the kind of

information scientists are interested in; in the model inaccurate parts cannot be, or cannot helpfully be, further analysed in terms of simplicity and idealisation (Rice 2019).

Finally, though simplification and idealisation are sometimes discussed as if they are distinct, it is worth noting that there is a way in which they can be understood to be closely related: Aristotelian idealisation, the stripping away of features, is in effect a kind of simplification.

2.4. Simple and complex models

The aim of this section is to contrast simple models with complex ones and in doing so arrive at a characterisation for simple scientific models which are the focus of this thesis. Complex models represent large numbers of causal or explanatorily relevant factors responsible for a phenomenon of interest. An example of a complex model is a state-of-the-art global circulation model (GCM).

Frigg et al. (2015) report that, in these models, the Earth system is understood to include atmosphere, ocean and other subsystems which are divided into grid cells. Climate processes are flows of certain quantities such as heat and vapour from one cell to the other and are characterised by the dynamical equations which form the dynamical core of the GCM. These equations of the model represent many different processes in the climate system and are intractable with analytical methods; a computer is needed to generate a solution. A reduction in complexity for example, by replacing an interactive ocean circulation with prescribed sea surface temperatures, would result in faster simulation time (Frigg et al. 2015). In many cases the aim with complex models seems to be to represent accurately the real factors that influence the outcome of interest in the target system.

At the other end of the spectrum are simple (or even ‘very simple’) models. They are characterised by the small number of causal or explanatory relevant factors that they represent. Examples of simple models often discussed in the literature include the frictionless plane, the

idealised pendulum for mechanics, the sun +1 model of the solar system, the billiard ball model of gas for physics, the one-to-one sex ratio for biology and the Schelling segregation model for the social sciences. A simple model typically is one that involves substantial Aristotelian idealisation, and perhaps Galilean idealisation as well. Different types of simple model are discussed in the next section.

2.5. Four types of simple models

My aim in this section is to give a brief discussion of four types of simple model. These are not mutually exclusive and exhaustive types; rather, they each highlight important features exemplified by at least some simple models. Table 1 at the end of the section displays the different features associated with these different types of simple model.

2.5.1. Simple models that are toy models

Toy models are simple models that are extremely simple and highly idealised (Reutlinger et al. 2018). Idealisation here can include both Aristotelian and Galilean idealisation. Reutlinger et al. (2018 pp.1072-1075) distinguish between two types of toy models: embedded and autonomous. Embedded toy models are both models of phenomena as well as models of the theory. Such models are embedded into an empirically well-confirmed theory. The characterisation of an embedded toy model is derived from a similar distinction in mathematics between a framework theory and models of a framework theory: A framework theory is a set of interpretive sentences including an abstract

calculus and general laws; models of a framework theory are structures of which the sentences are true (paraphrased closely from Reutlinger et al. 2018 p.1073, citing Frigg and Hartmann 2012 section 1.3).

An example of an embedded model is the sun+1 planet model of the solar system. Here, Newtonian mechanics is the framework theory, and the phenomenon is the orbit of the planet around the sun. The framework theory plus model assumptions may consist in a system of the sun +1 planet, used to derive the motion of a single planet around the sun. The sun +1 planet model is a structure of which Newtonian mechanics is true. The model is simplified and idealised in that it ignores all other planets and stellar objects as well as forces other than gravity (paraphrased from Reutlinger et al. 2018 p.1074).

Autonomous toy models, in contrast to embedded models, are not embedded in a theory but are only models of phenomena. Reutlinger et al. (2018 p.1075) identify Schelling's model of segregation (1971) as an exemplary autonomous toy model. Racial segregation is the target phenomenon for the model, which might be instantiated in real world cities. The simplifying assumptions that the model makes are firstly, there are only two kinds of agents, black-and-white, who live in an extremely simple environment, i.e., a two-dimensional grid; secondly, that the agents are randomly distributed; and thirdly, that they interact using the simple behavioural rule that each agent moves to an empty space in her neighbourhood if less than 30% of her neighbours do not have her colour (paraphrased from Reutlinger et al. 2018 p.1077). Iterating the behavioural rule, given the assumptions, produces segregation even though the agent has only a weak preference for same-coloured neighbours. The model shows how racial segregation can evolve, even without strong preferences to be surrounded by members of one's own race.

The autonomous toy model is idealised in both Galilean and Aristotelian senses in several respects. For instance, each agent is assumed to know how many agents of each colour live in her neighbourhood which is an instance of Galilean idealisation. Likewise, every agent can move if she is

dissatisfied with the colour of her neighbour. Aristotelian idealisation occurs insofar as, the model leaves out various factors that would not in reality be thought to make a difference, such as whether the people have dishwashers at home, etc. The autonomous toy model is derived or constructed from reflection on the target phenomena; no theory is directly involved in its construction.

2.5.2. Simple models that are epistemic surrogates

Though also focusing on highly idealised and simplified models, Grune-Yanoff (2009) understands such models and their targets rather differently - as 'epistemic surrogates'. According to Grune-Yanoff, this type of model, which he calls a 'minimal model', does not contain any causally or explanatorily relevant factors of particulars at all. Rather, the explanatory factors are tied to types, without direct reference to instantiation of a real-world phenomenon. The model instead has as its immediate object of reference a concrete imaginary situation. The eventual object of interest may just be an abstract feature of a class of phenomena (Grune-Yanoff 2009).

Grune-Yanoff gives the example of economists investigating equilibrium models: they want to learn about real equilibrium states that are not confined to the models they study, though they for the most part do not isolate a class of concrete situations that exhibit such a feature. It is important to note that Grune-Yanoff starts with a surrogate relation to the target and not a simplification of the target. So, the model is a surrogate for various real-world situations, rather than a simplified representation of some particular one. It is perhaps plausible, though, to understand such models as involving extreme Aristotelian idealisation of a range of real-world targets.

Grune-Yanoff also highlights the dynamical aspect of these simple models. Such a model can be manipulated in various ways to achieve results - for example by resetting a parameter value in a modelling assumption and running a simulation or deriving results by hand. Also important is the interpretation of the model, since a formal structure can generate different results depending on its interpretation, for instance, in the case of game theoretic models. Model results are therefore generated by a formal structure and its interpretation. However, on the one hand there is the interpretation of the model that is distinguished and separate from, on the other hand, the system that the modellers eventually learn from, which is said to describe a parallel world. Such descriptions do not describe a real-world target but can describe an imaginary world that is a useful epistemic surrogate, insofar as it describes familiar features.

Interestingly, Grune-Yanoff ‘s detailed example is the Schelling model of segregation, introduced above in connection with the toy model view. As Grune-Yanoff sees it, Schelling used coins and a chequerboard as an epistemic surrogate for real phenomena. He used the patterns on the chequerboard to represent neighbourhoods and the coins to represent real people but not any real particular neighbourhood or people. Schelling’s findings (as noted above) were supposed to teach or show us how racial segregation could evolve in real life cities even if the agent had no strong racial preference. But it is important to note that the labels he gave calling coins neighbours and distributing preferences to them, bore no significant resemblances to concrete real-world situations. Even though for example there was some resemblance to real features of real neighbourhoods in the configuration of coins on a chequerboard, he did not justify the model’s specification with reference to real-world situations. The model functions as an epistemic surrogate, on Grune-Yanoff’s view. Reutlinger et al. (2018), however, apparently interpret the Schelling model as having a closer representational relationship to real cities. The Schelling model seems malleable enough to cope with this difference of interpretation – it may well be understood somewhat differently by different users who use it to perform different functions.

2.5.3. Simple models that are minimal models

A different type of simple model is what will here be called a minimal model, following Batterman 2009; Batterman and Rice 2014. This type of model explains patterns of macroscopic behaviour across systems that are heterogeneous at smaller scales (Batterman and Rice 2014), meaning that the details of the phenomenon of interest do not matter for a proper characterisation of that phenomenon. [1]

According to Batterman (2009), nature presents us with repeatable phenomena, and we ask what is responsible for that repetition. It is the repetition that captures our attention and becomes a dominant feature. The dominant feature along with the need to idealise makes us understand that the dominant feature is necessary to find an adequate representation for that feature; therefore, idealization is necessary, for representing[2] a phenomenon of interest. Idealisation here, turns out to be a means of focusing on those features which are constitutive of the repeatability of the phenomenon of interest.[3] The kinds of phenomena that minimal models apply to range from biology to physics. For physics there are the examples of modelling shocks, breaking and formation of

[1] It is useful to note that while I am using the term 'minimal model' as Batterman does, that there are other characterisations of minimal models in the literature. See e.g. Weisberg 2007 , and recall that Grune- Yanoff's type of model is commonly referred to as a 'minimal model'.
[2] Batterman changes his mind as to whether representation is the correct term to use in this context.
[3] We can see here that idealisation and simplification are not separate and furthermore idealisation is not separate from the representation of the phenomenon. Rice's holistic distortion view (2019) fits Batterman's conception.

droplets (Batterman 2009) and for biology there is the example of Fisher's one-to-one sex ratio model (Batterman and Rice 2014).

Batterman (2009 p.430) cites Goldenfeld (1994), articulating two opposite views about the way models are used. Firstly, there is the traditional view on which a model is a faithful representation of the physical system of interest and includes as many details as possible. The aim here is that as technology increases, what once was an idealised and simple model is now made more complex. In this way it is like Galilean idealisation as described above. Moreover, the eventual aim for a traditional model is for a convergence between model causes and the behaviours of real factors that influence the outcome of interest. The model at first is intractable, and idealisation is used to make the model more solvable, but as more sophisticated technology evolves with more powerful computers the once insolvable equation(s) can now be solved so we can then add back the details into the model. Secondly, in contrast, there is the minimal model, or the non-traditional view which finds virtue where the traditional view finds vice: not to let details get in the way. With minimal models, the aim is to capture some essential elements of the phenomenon (Batterman and Rice 2014). Batterman and Rice (2014) use the example of the of the observed one-to-one sex ratios in natural populations as target or a phenomenon of interest for a minimal model. The approximate one-to-one sex ratio is an instance of universal (or at least broadly instantiated) behaviour across heterogeneous system--- more specifically 'systems' that are biological populations with two sexes (via sexual reproduction). The model they discuss is Fisher's (1930) sex ratio model, which is a simple equilibrium model. It can be used to show why a roughly 1:1 sex ratio commonly emerges.

Minimal models are idealised by simplification, here idealisation and simplification are synonymous, so the details do not get in the way to characterise a phenomenon of interest. Minimal models can be thought to differ from typical toy models in that they do not contain Galilean idealisations, in the sense that there is no interest in removing the idealisations. Furthermore, simplifications or idealisations are given a holistic treatment for minimal models--- they are

considered as working together with the rest of the model, rather than being viewed in a piecemeal fashion (Rice 2019).[4]

2.5.4. Target-less simple models

Luczak (2017) identifies a fourth interesting type of simple model, which is target-less models. These models have various functions, including teaching, elucidating theoretical ideas, and exploring compatibility of concepts. A key feature of such target-less models is that their relationship to some other system of interest (whether theory or real system) is one of similarity or analogy, rather than representation.

Examples of such models given by Luczak include the Kac ring, the Ising model, the Baker transformation, and the Arnold cat map. The Kac ring, for instance was first introduced by Mark Kac in 1959 to elucidate statistical mechanical treatment of irreversible phenomena. The ring model is a simple and solvable model, and it is made up of n-sites arranged around in a circle in a one-dimensional periodic lattice. Sites are joined to their neighbours by an edge and several of the edges have a marker. Each site has a black or white ball. The ball and markers are like molecules that make up the gas but are not intended to represent them. The system evolves on a discrete set of ticks: each ball moves clockwise to its nearest neighbour. When the ball passes the marker, it changes colour, and this is analogous to (but again not intended to represent) changes in velocity of collisions

[4] However, Reutlinger et al. 2018 say that though minimal models may be distinct from the toy model, minimal models are better characterised by the toy model approach. I believe that minimal models deserve their own category because for the most part minimal models have only non- traditional idealisation by taking limits, and simplification by non-dimensionalising the equation.

of gas molecules. Such a model can be used, for instance, to elucidate the concept of reversibility in a pedagogical setting (Luczak 2017 pp.3-4) as I discuss in chapter 6.

2.6. **Conclusions**

After characterising scientific models and then simple scientific models as a subclass of scientific models, I explored several types of simple scientific models. These were: highly idealised and very simple toy models; simple models that perform an epistemic surrogate function; minimal models that aim to not let the details get in the way; and target-less models that do not represent systems of interest but relate to them by analogy or similarity. Different characteristics of these types of simple model are noted in table 1 below.

Table 1. A summary of the characteristics of different types of simple models.

MODEL CHARACTERISTICS	TOY MODEL	TARGET-LESS MODEL	EPISTEMIC SURROGATE	MINIMAL MODEL
Aristotelian idealisation				

Galilean idealisation				
Real targets				
Imaginary targets				
No targets				
Embedded models				
Models of phenomena				

Key:

= TARGET-LESS MODEL

= TOY MODEL

= MINIMAL MODEL

= EPISTEMIC SURROGATE

3. What is understanding?

3.1 introduction

The next step in the investigation of the value of simple models for understanding is to get clearer on what is meant by understanding. This is the topic of this chapter.

Historically, understanding was seen by philosophers of science to be the purely psychological notion attached to explanation. Understanding was the Eureka moment and deemed irrelevant (to philosophy of science) and subjective. Explanation, which was objective, was the true aim for science (de Regt 2013). However, over the last decade in the philosophy of science, there has been a resurgence of interest in understanding. This body of work investigates a number of issues: the question of factivity, or whether understanding is related to truth; the nature of the grasping aspect of understanding; the relation between explanation and understanding; and what types of understanding there are. After surveying some of this work in this chapter, I accept a set of claims about understanding, as outlined in the conclusion of the chapter, but I do not argue that any particular account is the correct one. Later chapters, discussing views on simple models and understanding, will often appeal to specific accounts. The discussion in this chapter will provide the context for those accounts, helping us to see how they fit into a broader ongoing philosophical discussion about understanding.

Section 3.2. presents several different kinds of understanding. Section 3.3. discusses factive and nonfactive views of understanding. Section 3.4. considers grasping as a condition for understanding. Finally, section 3.5. discusses explanation as a condition for understanding. While, as noted above, I will not adopt one account as the 'correct' view of understanding, it is important to outline a range of views in the literature to provide context for the discussion in later chapters of how simple models can facilitate understanding.

3.2. Kinds of understanding

Philosophers have distinguished various kinds of understanding: understanding "that", understanding "why", understanding "how", understanding "with" and objectual understanding. There are often arguments to the effect that understanding comprises just one of these types (Khalifa 2012), however I will not argue for a particular type of understanding. Rather in this section I will provide a brief taxonomy and overview of the kinds of understanding that are possible.

The first kind of understanding is propositional understanding or *understanding-that X*---, for instance, I understand that the cat is on the mat. On some views this is closely related to understanding-why e.g., propositional understanding is knowing why a proposition is true (Strevens 2013). Here, knowing a proposition is true is not enough to understand it; for instance, hearing from a reliable source that DNA contains mitochondria is enough to *know-that* it is the case but not enough to *understand-that* it is the case. Understanding that it is the case might, for instance, require knowledge of causes, which is a kind of knowing why (Elgin 2017). However, I paraphrase Strevens (2013 p.511) as asserting that the cat being on the mat is a separate issue to why the cat is on the mat. We can know that the cat is on the mat without knowing how it got there; how it got

there would give us a sort of understanding-why. Understanding-that may nevertheless be closely linked to understanding-why.

Understanding-why can also be given a separate analysis. This kind of understanding is in the form: I understand-why X. So, for example, I can understand-why the cat is on the mat as opposed to understanding-that the cat is on the mat. This implies that understanding-why is a different matter to understanding-that. However, one way in which understanding-that may be closely related to understanding-why is when understanding-why is cashed out in terms of grasping propositions which is a matter of understanding-that (see Strevens's simple view discussed later). Understanding-why is often thought to involve explanation.[5]

Thirdly there is understanding-how: This is closely related to understanding "that" and "why". It can however be given a separate analysis to understanding-that, because here, understanding-how might not be propositional, it might be tacit. For instance, a mechanic may understand how complex machinery can work without being able to articulate it propositionally; conversely, in other cases, understanding-how may be translated into a list of understanding-that.

Fourthly there is understanding-with. The object of this kind of understanding is a phenomenon or a state of affairs which can be understood via a theory. To understand with, therefore, is to use the theory to explain a wide range of phenomena. The wider the range of phenomena you can explain, the greater your understanding.

Finally, there is objectual understanding which is understanding of a topic, subject matter, or a body of information (Elgin 2017). Here coherence linking bits of information together into a body of

[5] Parker (2014) in the context of climate science, says that understanding why is closely linked to explanation, (see also Khalifa and Stevens above), so that you gain this type of understanding when you grasp a correct explanation of the phenomenon. She contrasts this with understanding a complex phenomenon or system. Though not made explicit this latter kind of understanding is a species of objectual understanding (see below). However, she does state that the two kinds of understanding may be co-dependent.

information is the crucial concept that results in an objectual understanding. However, coherence alone is not enough: Ptolemaic astronomers had a coherent account of celestial bodies, but we do not say that they understood the motions of the celestial bodies (paraphrased from Elgin 2017). We therefore need some relation to the relevant facts. An obvious candidate is truth, such that the propositions of a body of knowledge must therefore be true. However, truth as a criterion is too strict if we want to allow that the sciences give objectual understanding of their subject given the pervasive use of idealisations. As a response we can replace truth with approximate truth. However, it is unclear how some false models can even be approximately true.

In response there are at least two views that depart from truth which do not defeat understanding: the first view says that only claims about the difference makers need be true (Strevens 2008). So, the non-interaction of the particles in the ideal gas law makes no difference to the law's applicability: It is in fact true that the ideal gas law holds for some domains in phase space, and where it is false, there is a story to be told (Elgin 2008 p.7). The second view is that truth is one amongst many [and it is not the most important] modes of accuracy. Models and idealisations share exemplified features of the subject matter and this will allow epistemic access to them, as discussed below (Elgin 2008, 2017).[6]

There is another (more holistic) version of objectual understanding emphasising that objectual understanding comes in degrees. This version assumes four conditions for objectual understanding (Baumberger and Brun 2016, Baumberger 2019) including a commitment condition, and a condition that answers to the facts (a rightness condition). There is also the condition of grasping and finally a justification condition. The content of these four conditions are elaborated upon in chapter 6 for a generalised theory of objectual understanding.

It is worth making clear how the object of understanding for objectual understanding differs from that for explanatory understanding. One can stipulate that the difference is only by degree

because, in explanatory understanding, the object of understanding is a state of affairs and in the case of objectual understanding, the object is merely a system of state of affairs or a complex state of affairs. So, authors such as Grimm (2011) argue that the purported difference between objectual and explanatory understanding is spurious.

However, Baumberger and Brun (2016) argue that this is a hasty conclusion because objectual and explanatory understanding are also distinguished in terms of how they are achieved. They follow Kvanvig (2009) who notes the similarities to isolate differences. Kvanvig maintains that both forms of understanding grasp structural relationships e.g. probabilistic, logical, and explanatory relationships between the items of information out of which the question of understanding arises. The main difference is in the kinds of structural relationships pertaining to each kind of understanding. So explanatory understanding is more restrictive with respect to the inherent structural relationships than objectual understanding. The canonical example of attributions of explanatory understanding reflect this in so far as: S understands why P, where S refers to a person and why P refers to an embedded question. In such cases explanations are identified with answers to "why" questions. In contrast to the above canonical example of attributions of explanatory understanding, we allow objectual understanding to incorporate explanatory relations when they exist. But when they do not exist, objectual understanding is achievable by grasping other structural relationships i.e. logical or probabilistic. So, we say that objectual understanding has a wider scope than explanatory understanding, though Kvanvig stipulates they may be compatible. However, if the target of understanding is a theory, the compatibility of explanatory and objectual understanding may not exist because the target of explanatory understanding is a phenomenon, and in objectual understanding, it can include things other than phenomena.

In the following chapters, I will assume that all these varieties of understanding are of interest when it comes to the question of how simple models can facilitate understanding. In other words, we are interested in how simple models can facilitate any one of these kinds of understanding. We

will see that Chapters 4 and 5 are concerned primarily with understanding-why and that chapter 6 is primarily concerned with objectual understanding.

3.3 Factive and non-factive understanding

In this section I provide an overview of different views of whether understanding is factive. I will discuss strict factivism, a weak factive view, and a non-factive view. Throughout, I will refer to the basis (P) for understanding, i.e., that in virtue of which we come to understand the object of understanding (X), and what we come to understand (U) about that object, (paraphrased from Khalifa et al. 2019 pp.346-347).

Strict factivism about understanding requires that all the beliefs and the way in which those beliefs have been arrived at, about a subject matter ---so this is (P) and (U) above--- are true. Such a position may be appealing because it preserves the intuition that understanding is a desirable epistemic "good". However, the limitations of strict factivism are abundant. Most importantly, it lies in tension with the fact that idealisation is an essential and pervasive aspect of scientific theorising (Rice 2015). Idealisations are false assumptions. Our best models have known false assumptions, and it seems that we think that at least some of these models sometimes give us genuine understanding. So strict factivism is too strong.

Weak Factivists about understanding take the view that it is not the case that all beliefs (U) and the way in which we arrive at those beliefs (P) have to be true (Elgin 2017). This would presumably incorporate the stronger thesis: that you get understanding of (X) resulting from (P) (which is true) (Khalifa et al. 2019). So weak factivism allows there to be understanding from P which is strictly true.

One version of the weak factivist view is presented by Elgin (2017)[7]. It holds that only the central beliefs of (P) need to be true. Central beliefs are the core beliefs in a belief system (P) and can be contrasted with more peripheral beliefs which are held more contingently. Such a weak factive view has an advantage over the strict factive view insofar as some idealisations are allowed; but they are peripheral and, in a sense, do not matter. It is also in more accord with scientific practice in so far as a single false belief or minor aspect of the subject matter being false would make no difference to the attribution of understanding. Furthermore, it is compatible with the idea that understanding comes in degrees: I can have better understanding of (X) than you if more of my central and peripheral beliefs in (P) are true.

However, there are limitations to this weak factive view too: Firstly, in virtue of what does a belief count as central? Secondly, weak factivism says that understanding cannot be attributed with a single central false belief. However, sometimes when it is especially clear where in our basis for understanding (P), we have one false central belief, but a lot of peripheral true beliefs we would still want to say that there is understanding. The weak factivist owes an explanation as to why this may not be the case.

While this weak factive view has some kind of truth condition or accuracy requirement in terms of central beliefs, Rice (2015) maintains a different weak factive position without the requirement that central beliefs of (P) be true. He subscribes to the view that understanding of phenomena requires that what one understands about the phenomenon (U) must be systematically integrated into a wider body of information about the phenomenon of interest (Rice 2015).[8] His contention is that scientific understanding requires only that *most* of what one believes to be contextually salient propositions (P) are true.

[7] Note that this is not her view. Her view is a non-factivist view, as discussed later.
[8] Rice concentrates on modal / possibly relationships of information rather than for instance causal, functional, logical or exemplar relationships.

Rice (2015) has a case-by-case approach that allows for a plurality of context sensitive ways for understanding to meet this weak factive requirement. The context establishes a typical why question with a contrast class--- a set of features that are relevant and irrelevant. He emphasises contra Strevens (2008) that accurate representation of difference-making features is not always needed for understanding. For instance, it may be enough that the model and the phenomena are in the same universality class, which entails that they will display similar macro-level behaviour. Such a model might be used to obtain true beliefs about what is possible, and then these beliefs are incorporated into a large network of by and large accurate information.

The last alternative is non-factive understanding. For the non-factivist, a radical departure from the truth does not necessarily prevent understanding, though this is perfectly compatible with there being understanding from a true basis too. One argument from the non-factivist is broadly this: if we take a closer look at the science, we shall see that there can be cases of non-factive understanding (Khalifa et al. 2019). For greater precision analytically, Khalifa defines non-factivism about understanding as allowing:

There exists some basis P and target of understanding X such that:

1. P is False
2. P provides understanding of X, and
3. The understanding of X resulting from either not accepting P or accepting a more accurate proposition instead of P is not better than the understanding provided by accepting P (Khalifa et al. 2019 p.346)

To motivate such a view, Khalifa et al. (2019) argue that understanding of the ideal gas law [X] via a simple derivation from clearly false assumptions (e.g., non -interacting particles) is *no worse* than the understanding provided by more sophisticated approaches (e.g., involving the Van der Waals forces, the virial expansion).

This is in opposition to the quasi-factivist, what Khalifa calls 'the weak factivist view', who contends that understanding increases as derivations are performed to arrive at more accurate representations that replace the ideal gas law (Khalifa 2019 and Mizrahi 2012).

Khalifa (2019) challenges this quasi-factivist position in several ways. For one thing he notes that the quasi-factivist begs the question when they simply assert that the more accurate explanations provide greater understanding; rather, this must be defended.

In addition when considering the ideal gas law versus the van der Waals derivation, he claims that there is a shift in the object of understanding from one derivation to the next; there are two objects of understanding, the object of the ideal gas law and the object of the van der Waals equation. Moreover, the basis for understanding one object is not without further qualification and does not provide the basis for understanding the other object. If there are two different objects with two different bases, this complicates the claim that our understanding increases from one derivation to the next (paraphrased from Khalifa et al. 2019 pp.355-356).

A deeper objection to the quasi-factivist, in the same vein, considers the virial expansion. The virial expansion is arbitrarily precise in terms of representation of particle interactions, making it more accurate than the ideal gas law. Also, the ideal gas law can be derived from the virial expansion, therefore one can assume that the virial expansion as more fundamental. However, Khalifa et al. (2019 p.357) points out that there are two problems. Firstly, the ideal gas law's derivation from the virial expansion also involves idealisations. So, at best the quasi-factivist in contemplating that the ideal gas law is derived from the virial expansion, only shows how one idealisation---- that of noninteracting particles---- piggybacks on another, more fundamental

idealisation: in this case the thermodynamic limit. So even if idealisation of non-interacting particles is a way station this would leave the idealisation of the thermodynamic limit as ineliminable.

Secondly it is not obvious that the virial derivation provides greater understanding than the ideal gas law. To start with, one might entertain that the virial derivation is deeper than the ideal gas law because it shows how other idealisations work. It says why particle interactions are relevant to the behaviour of ideal gases. But it does not show us that particle interactions are irrelevant (Strevens 2013).

Another view of non-factive understanding comes from Elgin (2009) who has a different approach than Khalifa. Elgin contends that science is not indifferent to the facts but relates to the facts in more complicated ways than the factivist admits to. On her view scientific understanding involves felicitous falsehoods. These falsehoods are not true, but they are also not defective by being false.

For example, gas molecules are not spherical, and if a model contains such spherical molecules, this could be a felicitous falsehood; a fiction; that helps someone gain insight into the nature of the target phenomenon. The explanatory power of felicitous falsehoods may be greater than truths e.g., in the case of thought experiments such as the gas laws whose pressure, temperature, and volume are interdependent, such experiments are in fact fictions (paraphrased from Elgin 2012 p.9). The thought experiments that are played out in our head tell us something about the behaviour of gases, telling us something more than truths would i.e. the *behaviour* of gases. How can this be the case? Where do we draw the line using fictions in science? To answer the question of the limits of fictions in science, Elgin appeals to Goodman's notion of exemplification. Successful models or thought experiments exemplify features that they share with their targets and refer to the targets by that exemplification (paraphrased from Elgin 2009 p.10).

What is exemplification? It is a relation of the sample to what it is a sample of (Goodman 1968, Elgin 2009 pp.10-11). A sample is an instance of something. Therefore, an exemplar is a symbol that

refers to some of the properties it instantiates. It is selective, for example, a teacher may use a student's work as an exemplar of the kind of essay she wants her students to write and depending on her use of the exemplar, she might exemplify a certain aspect of the essay, for instance, its form. Exemplars are also involved in dual representational relationships: an exemplar refers to a property it instantiates and to the extension of that property. Elgin stipulates that a well-chosen exemplar gives epistemic access to its own properties and to a wider class. But as an exemplar is a symbol, it therefore needs to be interpreted and because its reference depends on context, so does its interpretation depend on circumstances (paraphrased from Elgin 2012 p.14).

Applied to science we can say that the conditions in the model cannot occur in nature, but the model has value in highlighting a factor in what occurs in nature enabling scientists to discern and play with its consequences in such a manner that they would not be able to, by any factual representation (paraphrased from Elgin 2012 p.16). For example, by interpreting the behaviour of actual gases as variations on the ideal gas, we understand the behaviour of actual gases better than if we had a realistic picture of the complex interactions among different sized and shaped actual molecules of gas, thereby making use of exemplification. So at least in such cases, simple models provide better understanding than more complex ones. The contention is that truth is not the only secure link between theories and the world. Exemplification can supply another strong tether to the world for the case of simple and highly idealised models that are pervasive in science.

The main drawback of a non-factive account is that it might be too permissive. Maybe there are cases involving radical falsehoods where we do not want to say there is understanding. Nevertheless, I will in the remaining chapters allow for at least some non-factive understanding. Strict factivism is clearly ruled out given that I am exploring how simple (idealised) models facilitate understanding. Sometimes, the simple models argued for in later chapters to facilitate understanding will meet the weak factivist standard and sometimes they will not. The important

thing to note is that we should not insist on weak factivity in all cases. This just is the non-factivist position.

3.4. The grasping condition

Some accounts of understanding involve a grasping condition. However, if understanding is non-factive, some accounts of grasping will not work. I will discuss one account of grasping that is only compatible with factivity, and one that will work for a non-factive view of understanding Strevens 2013). I will also present a modified account of Strevens's view given by Reutlinger et al. (2018).

Strevens (2013 p.511) asks the question: what is it to grasp that a certain state of affairs obtain? He says to grasp a certain state of affairs obtain is to understand that a certain state of affairs obtains. The grasping here is a kind of understanding, but he asserts that this kind of grasping is not synonymous with the understanding used in a theory of understanding (paraphrased from Strevens 2013 p.511). For grasping, Strevens appeals to a distinction between "that" and "why" kinds of understanding (see sections 3.1 above); to understand *that* the cat on the mat is different than to understand *why* the cat is on the mat. "That" kind of understanding comprises grasping and the "why" kind of understanding is the target of a theory of understanding such as the simple view discussed below (paraphrased from Strevens 2013 p.511).

By asserting that understanding comprises grasping, we are now able to enquire into the nature of understanding that or grasping per se. One contention is that knowledge or even something more than knowledge is what defines grasping. Strevens (2013) stipulates grasping is more than knowledge, that is, it is over and above knowledge. For example, someone can know the chemical properties of water with only a little understanding of chemistry; they can know that water is H2O, but we say that they don't, just by asserting water is H2O, thereby grasp the state of affairs that hold

relevant to the above (grasping is more than having knowledge) sense of understanding. This is shown more clearly by Strevens (2013 p.511) with another example, one that involves understanding the phenomena explained by Newton's second law. Suppose that we take Newton's theory of gravity to be correct and that someone knows its tenets and knows, given background conditions that they entail Kepler's laws. So, they know the propositions that comprise the correct explanation of Kepler's laws and that these propositions stand in the correct relation (for example using Hempel's DN account of explanation, the propositions stand in a deductive argument). However, we can still say that the propositions are not grasped sufficiently for understanding. For instance, even though they know that the second law is true, they may grasp only a little of its content so they cannot fully understand the phenomena explained by the second law. For Strevens then, grasping needs a more intimate acquaintance with the structure of explanation than is present with most accounts of knowledge. So, in answer to the question what is grasping, Strevens (2013 pp.511-513) stipulates that grasp is the fundamental [important] relation between the mind and the world--- a kind of direct apprehension.

However recently, Strevens (2020), unpacks grasping slightly differently, in the following way: it is recognition. Grasping a property, for example is equated with the ability to *recognise instances of* that property. Recognition is an inference to a conclusion of the form "this is one of those". Therefore, recognition is a capacity or an ability to place things in a domain. Recognitional capacities must latch onto something in the world. A recogniser must be able to, as it were, "recognise in the wild". Further, grasp comes in degrees, and the wider the range of circumstances you can recognise, the firmer the grasp.

Minimally accurate (false) models can also be treated by Strevens's theory of grasp. The idea is that you have a grasp to the degree that you would recognise these things *if* you "saw" them. So, grasping a false model is just recognising that model, as if the sentences of the "world" described by

that model *would* have been true. This tests our recognitional powers, and so the model "world" is the natural environs for the minimally accurate model.

Therefore, grasp has two dimensions---- accuracy and scope--- and is also one-sided in the sense that, apart from recognition, no further skills are needed, i.e. there is no need for an output in the sense of manipulation and exploitation of what is grasped, e.g. by building a model and using it. Because grasp is only a first step towards the totality of understanding that is possible regarding a property or fact, it is an entry point to understanding.

Given that grasping is in terms of understanding-that, it seems that grasping must be factive. However, this condition may not work for accounts of understanding-why that are of interest to the present discussion of simple models: we want an account of grasping for a non-factive account of understanding. This is accommodated by Strevens (2013) who supposes that grasping is made up of two components:

[1] the purely psychological or narrow component, and

[2] the obtaining of the grasped state of affairs (Strevens 2013 p.512).

For the non-factivist, grasping in terms of only the psychological component implies that the cat really does not have to be on the mat if we are to grasp that "the cat is on the mat". So for the psychological component of grasping, it should persist in one's mind even if an evil demon, at the moment of your grasping that the cat is on the mat, were to remove the cat while still maintaining in your mind the appearance of the cat being on the mat. So, grasping for the non-factivist, is, then, the internal psychological state.

Reutlinger et al. (2018) refine Strevens's approach by naturalising grasp. They interpret Strevens as saying that grasping is a fundamental relation between mind and the world, but they also see grasping as philosophically primitive though not scientifically primitive. What does this mean? It means that grasping has a publicly accessible component, the philosophical component, and a subjective component which takes place in the individual's mind. What the subjective component of grasping is, is a matter for science, rather cognitive science, not philosophy.

In my analysis in the forthcoming chapters, I will assume that grasping is involved in the understanding gained from simple models. At the very least, it involves a narrow psychological component of grasp (Strevens 2013), as articulated above.

3.5. Explanation and understanding

This section investigates how explanations are related to understanding. Most accounts of understanding have something to say about the relation between understanding and explanation (Khalifa 2012, 2013, Strevens 2013, Lipton 2009 and De Regt 2017).

It can be stipulated for example that explaining "why" and understanding "why" are closely related: explanations give answers to "why" questions and to understand is to have those answers (Lipton 2009). Explanation here is propositional and explicit, and it is also in the form of an argument. Specification of the structure of explanation alone, however, does not allow for identification with understanding. Why should things that have that structure be identical to understanding? Here understanding is a cognitive achievement, it is more than knowledge that a phenomenon occurs. More precisely, understanding is the cognitive benefits that explanation provides knowledge of causes, of necessity, of possibility and unification. These cognitive benefits

give information about the causes, demonstrate that the phenomenon had to occur, how the phenomenon occurred and lastly how it fits into a broader pattern (paraphrased from Lipton 2009 pp.43-44). In this way we distinguish explanations per se from understanding by identifying understanding with the cognitive benefits of explanation. This implies that understanding can in principle come from ways that do not involve explanation.

An advantage of a conception of understanding that does not require explanation is shown by the case of understanding-how (see the first section on types of understanding); here one might argue, understanding flows from abilities and not from explanations (Lipton 2009). However even for understanding-why as a form of propositional knowledge, according to Lipton, we can get understanding without actual explanation, where abilities support conventional forms of understanding such as knowledge of causes, unification, and modal knowledge. So, on this view understanding is equated with the cognitive benefits that the explanation provides as opposed to the explanation itself (paraphrased from Lipton 2009 p.44).

Khalifa (2013) interprets Lipton in a different light, one that he sees as consistent with Lipton's analysis, namely, that explanation may be the ideal of understanding. On this view, the understanding that explanation provides is the yardstick we use to measure other understanding. Khalifa (2013) contends that for all the examples given by Lipton, there exists an explanation that provides a better understanding of the phenomenon. This position may be characterised as explanatory idealism which entails that for every instance of non-explanatory understanding, an explanation exists which provides greater understanding (paraphrased from Khalifa 2013).

On some views that link explanation and understanding, reminiscent of the factivity discussion above, the explanation involved in understanding must be in some sense correct. Correctness, however, may not entail literal truth of assumptions of a model involved in explanation. Strevens (2013 p.512), for example appeals to the idea of a translational manual which picks out the relevant explanatory content. An idealised model that assumes "all F's are G's" translates to and therefore

has the explanatory content that “almost all F’s are G” or alternatively that “in conditions C, all F’s are G”, where C is determined by the context of production, for example, this could be done by the intentions of the explainer (paraphrased closely from Strevens 2013 p.512). What matters is that that this translated explanatory content be true.

Another relevant argument given by De Regt (2017) for the central role of explanation in understanding is also from the practice of science. De Regt is concerned with a “why” kind of understanding (section 3.2. above), so that we gain scientific understanding by answering “why” questions and such “why” questions necessitate explanations. Understanding is a product of explanation, so understanding comes from having an adequate explanation of the phenomena. Understanding, here, is pragmatic and thus context dependent though this does not entail subjectivity.

De Regt holds that scientists need intelligible theories to have scientific understanding of phenomena. Intelligible theories are involved in model construction by which scientists derive the explanation of phenomena on the basis of relevant theory. De Regt formulates a criterion for understanding phenomena:

CUP: A phenomenon P is understood scientifically if and only if there is an explanation of P that is based on intelligible theory T and conforms to the basic values of empirical adequacy and internal consistency (De Regt 2017 p.92).

The basic idea is that we understand phenomena via intelligible theories. CUP forms the basis of a theory of scientific understanding that is implicitly pragmatic because of the criterion of intelligibility. Intelligibility is pragmatic because it depends on the qualities of theory and on the scientists involved; whether scientists see a theory as intelligible depends for example on their skills

and background knowledge. CUP also implies that the phenomenon is understandable in virtue of the epistemic framework accepted by the community.

To qualify intelligibility, De Regt proposes a further criterion that is the criterion for intelligibility of theories (CIT). The definition for intelligibility according to De Regt is that it is a value that scientists attribute to the cluster of qualities of a theory that facilitate the use of that theory. De Regt (2017) formulates CIT by noting that a scientific theory is intelligible for scientists, only in a particular context, if they can recognise the qualitative characteristic consequences of the theory without performing exact calculations.

In the following chapters, I will assume that understanding can be intimately connected with explanation, as an ideal if nothing else. I think the intuitions of Lipton (2009) are successfully captured by Khalifa (2019), De Regt (2017) and Strevens (2013) are on the right track. We can have different kinds of understanding without recourse to explanation, yet explanation at the very least figures in understanding insofar as explanatory understanding remains the ideal.

3.6. Conclusions

In this section I recap what I will assume about understanding in the remainder of the thesis as I explore how simple models can facilitate understanding. Firstly, there are many kinds of understanding: understanding “that”, “how”, “why”, “with” and objectual understanding. A model facilitates understanding if it facilitates any one of these types, but my analysis will focus on understanding “why” and objectual understanding. Secondly, while understanding will often involve at least weak factivity, this will not be strictly required, and, given this, I assume the non-factivist

position which allows there can at least sometimes be understanding from information that is largely or strictly false. Thirdly, at the very least, grasping in the narrow sense (Strevens 2013) --- involving an “as-if” sensation—will be involved. Finally, it will be assumed that explanation is often closely linked to understanding. Understanding does not have to involve explanation (Lipton 2009), but consistently, explanation can be the ideal of understanding (Khalifa 2012).

4. Simple models facilitate understanding- - the case for how-possibly and how-actually explanations

4.1. Introduction

In this chapter I will give a first answer to the question: how do simple models facilitate understanding? The answer is: by providing how-actually and how-possibly explanations. Section 4.1. introduces the chapter. Section 4.2. looks at how-actually explanations. Section 4.3. discusses how-possibly explanations. Finally, I will elaborate on the link between explanation and understanding as is appropriate to this chapter in section 4.4.; I will mainly consider Reutlinger et al.'s (2018) refined simple view of understanding. One of the points in section 4.5. is that there is no conclusive way to draw the line between how-actually and how-possibly explanations because of the variety of views that suggest agreement on how to distinguish between how-possibly and how-actually explanations may be difficult.

To begin, I note that various scholars draw the distinction between how-actually and how-possibly explanations in different ways: for instance, Dray (1950) says that how-possibly and how-actually explanations are different kinds of explanations while Brandon (1990) puts the distinction on a continuum, and Bokulich (2014) further has incremental evidence differentiating the two concepts. However, in the remainder of the chapter I will be focusing only on some of these views, particularly the how-actually and how-possibly explanations of Reutlinger et al. (2018), which relate closely to those of Bokulich (2014) and to the how-not-possibly hypotheses of Grune-Yanoff (2009).

4.2. The case for how-actually explanation

In this section I will discuss how simple models facilitate understanding by providing how-actually explanations. I will then follow Reutlinger et al. (2018 pp.1089-1083) in noting that it is not the case that all toy models give how-actually explanations; some toy models do not provide how actually explanations and therefore also do not provide how-actually understanding. Also, I will outline another notion of how-actually explanation by Bokulich (2014) who extends Reutlinger et al's (2018) conception of how-actually explanation.

4.2.1. Reutlinger et al's view

The question is how do simple models, in this case, toy models, facilitate understanding? One answer is: by providing how-actually explanations. To show how this works, I need to recap the idea of embedded toy models. Embedded toy models are models of well a confirmed framework theory, and they are simple and idealised models of phenomena (Reutlinger et al. 2018). What matters here is that the sentences of a well confirmed theory are approximately true of these models [via interpretation].

Therefore, according to Reutlinger et al. (2018), embedded toy models facilitate how actually understanding by yielding how actually explanations given these two conditions:

A. The well confirmed embedded framework theory T permits an interpretation and justification of the idealisations of M and

B. This interpretation and justification are compatible with the veridicality condition.

(closely paraphrased from Reutlinger et al. 2018 p.1086).

Reutlinger et al. (2018) argue that above the two conditions, A and B, are sufficient for how-actually understanding. They do this with an illustration from a model of Newtonian mechanics: the sun+1 planet model. This model is an embedded toy model because it is embedded in an empirically well confirmed framework theory, in this case Newtonian mechanics. It is a toy model because it is simple in describing a system with only two bodies, and idealised because it disregards other planets' influence and is a model of the target. So, the capacity of the sun+1 planet model to facilitate how-actually understanding depends on:

1. The real possibility to interpret and justify idealisations by, for instance, following McMullin's strategy for idealisations. So, the model is at least approximately true of the target i.e. the real orbit of the earth around the sun. These idealisations are justified pragmatically in other words they ease the calculation of the orbit.
2. That McMullin's idealising strategy is compatible with the veridicality condition; not only is the model approximately true, but its relation to truth is also further supplemented with de-idealisation i.e., that corrections to the idealisation can be put back into the model at a later stage, thus making it true of the target system (closely paraphrased from Reutlinger et al. p. 1087)

Newtonian mechanics has the resources and provides guidelines that would allow such a de-idealisation in the model to happen. Such a de-idealised model would include the influence of other planets and moons to make predictions and have greater accuracy than the toy model at hand. So, it can be said that in this way the McMullin strategy allows us to meet the veridicality condition.[9] It meets the veridicality condition just in case the framework theory has the resources to interpret and justify the idealisations made in an embedded toy model.[10]

4.2.2. Bokulich's 'how-actually' explanation

There is another way to get how actually-explanation other than via being embedded in a theory (Reutlinger et al. 2018), suggested by Bokulich (2014), though it can be said that her thesis is only an elaboration of Reutlinger et al.'s 2018 conception of how-actually and how possibly-explanation. Her thesis is that if one pays adequate attention to the different contexts in which these explanations are being deployed and the different levels of abstraction at which the explanandum phenomenon can be framed, we can make the distinction between how-actually and how-possibly explanations (Bokulich 2014). As my concern is with how-actually explanations, I will use her example of the Tiger Bush to illustrate how-actually explanations are generated.

[9] It is important to note that McMullin's strategy is one amongst many ways to interpret and justify idealisation. Reutlinger et al. think that the case of minimalist idealisation i.e. the idealisations involved in, for instance, Batterman's minimal models are compatible with the kind of de- idealisation, see for instance Maaki's de-isolation. However, this characterisation of minimal models may be contentious see for instance my characterisation minimal models in chapter 2, (cf. Batterman and Rice 2016, Batterman 2009 as well as the next chapter 5), on how minimal models facilitate scientific understanding. However, Reutlinger et al. 2018 say that the accounts of minimal idealisation such as the Strevens's (2008) minimalism do not give warrant to the claim that all autonomous toy models generate how actually understanding.

[10] Fisher's sex ratio model is also considered embedded toy model by Reutlinger et al.2018, one that presumably facilitate how actually explanation. The embedding framework theory would be Darwinian natural selection.

The case study Bokulich (2014) uses does draw conclusions about how-actually and how-possibly explanations of the Tiger Bush, which is a banding of vegetation in semi-arid regions. These vegetation stripes are called the Tiger Bush because they resemble the stripes on a tiger. The Tiger Bush is not generated by local heterogeneities in topological or soil considerations and so a question concerning the Tiger Bush is: what is responsible for the formation and maintenance of these stripes? Bokulich (2014 p.326) notes that it is likely to be a self-organised phenomenon resulting from intrinsic vegetation dynamics. The primary tools that scientists use to explain the emergence of vegetation patterns are computer simulations involving various mathematical models that are highly idealised. In this way the explanatory model of the Tiger Bush is a simple model. The question now is how such a model of the Tiger Bush provides a how-actually explanation?

To illustrate an answer, I will only concentrate on the first step of a tree like structure of explanatory models, increasing in specificity for the case of the Tiger Bush. The first step in the case of the Tiger Bush is to know that it is likely to be a self-organised phenomenon, with no external drivers, given that there are no heterogeneities in topography or soil in the formation of stripes. However, there is a prior split, before the assumption of self-organisation, that is, between the self-organisation model and the external forcing model. The decision for the self-organisation model as the model operating in nature in the Tiger Bush case comes from consensus made from field data revealing no variations in topography, soil, and other external drivers. This consensus among the scientific community makes it a how-actually explanation because of agreement that it is the true explanation, the one operating in nature. Prior to this agreement, both the external forcing model and the self-organisation model are how possibly explanations.

It is important to note that on the Bokulich (2014) picture, the more abstract the how-possibly explanation is, the easier it is to establish that explanation as a how-actually one. This is because there is sufficient generality of the claim such that it may describe a real mechanism in nature accurately. So, the level of explanation of the phenomenon i.e. its level of abstraction is important,

and the level of abstraction corresponds to different explanatory contexts. The different explanatory contexts have as their origin, relevant contrast classes of possible explanations. What makes it a how-actually explanation is evidence which underwrites consensus among the scientific community. This is a different concept to the how-actually-explanation concept of Reutlinger et al. (2018) who present how-possibly explanations as partial how-actually-explanations (see below). For Bokulich (2014), how-actually explanations are on opposite poles to the how-possibly explanations, on a continuous spectrum, so what starts as a how-possibly explanation becomes a how-actually explanation, given evidence and consensus within the scientific community. This interaction between how-possibly how-actually explanations is how Bokulich (2014) extends the Reutlinger et al. (2018) picture of how-actually explanations.

4.3. The case for how-possibly explanations

In this section I will answer the question of how simple models facilitate understanding slightly differently by providing how-possibly explanations. Firstly, I will expand on the thesis in the previous section (which stipulates that not all autonomous toy models that seem to be explanatory yield how-actually explanation) following Reutlinger et al. (2018): that some autonomous models yield how-possibly explanation. Secondly, I will elaborate on the Bokulich (2014) conception of how possibly-explanation and show how it extends the Reutlinger et al.'s (2018) view of how-possibly explanation.

4.3.1. Reutlinger et al. on how-possibly

Another way to answer the question of how simple models facilitate understanding is by providing how-possibly explanations (Reutlinger et al. 2018 pp.1093-1095). To recap, the notion of autonomous toy models: they are toy models that are not embedded in a well confirmed framework theory. However, they are, like embedded toy models, simple and idealised. Reutlinger et al. (2018 pp.1075-1077) cite the example of Schelling's model of racial segregation as an example of an autonomous toy model: Racial segregation in real world cities is imagined, contingently instantiated in the model. The model, however, has simple assumptions e.g. agents are randomly distributed, and they interact by a simple behavioural rule. So, given random distribution and the iteration of the behavioural rule, running the model, we get the emergence of segregation. The model is simplified and idealised to the extent that it does not accurately represent the preferences of real-life inhabitants for instance in Chicago's highly segregated neighbourhoods in the 1960's (paraphrased from Reutlinger et al. 2018 p.1075).

The reason for achieving only how-possibly explanations in the case of autonomous toy models, is that the embedding theory is simply not there to guide the de-idealisation to comply with veridicality. There is no available resource for guiding the interpretation and justification of the idealisations. Reutlinger et al. (2018) suggest that even if, for instance, the conception of the minimalist interpretation of idealisations [see for instance Strevens (2008) which complies with veridicality] were to be applied successfully to these autonomous toy models, this would not generate a how-actually explanation. For example, they contend, that it is not the case in Schelling's model that the idealised assumptions refer to explanatorily irrelevant factors.

This is how the Schelling model's idealising assumptions do not refer to explanatory irrelevant factors according to Reutlinger et al. (2018 p.1076): firstly, there is the fact that each agent is assumed to know the colour of his or her neighbour and a number of coloured inhabitants in his or her neighbourhood. Secondly, every agent is assumed to be able to move if she or he is dissatisfied with the colour of his or her neighbour. Thirdly, social, and economic factors are not considered to

make a difference. Fourthly, the model assumes random distribution of its inhabitants, for instance, in the particular city that the model is about. Each one of these factors is not explanatorily irrelevant in the sense that they are important or ineliminable to the functioning of the model as a model that can show the evolution of racial segregation with a minimal behavioural segregation rule.

Reutlinger et al.'s (2018) refined simple view of understanding applies directly to how-possibly explanations. This says that:

Scientist S understands phenomenon P via model M and in context C if and only if the following condition holds:

Scientist S has a how-possibly understanding of the phenomenon P via model M and in context C if and only if the model M provides a how possibly explanation of P and S grasps M.

(Reutlinger et al. 2018 p.1086).

Reutlinger et al.'s Schelling model explains how it is possible that racial segregation occurs. The model does not explain how it actually occurs but displays one possible route in presumably many routes in which it could occur. This model provides only a potential explanation of the general pattern (Reutlinger et al. 2018) and the general pattern is the emergence of racial segregation and this pattern is instantiated in real world cities like Chicago in the 1960's. More importantly such models do not necessarily provide correct identification of explanatory relevant factors, so a further,

related question is: why is it those scientists are interested in how-possibly explanation if they gain less, i.e. merely potential understanding, than with how-actually explanations? So, in other words what are the merits of how-possibly explanations? The answer that Reutlinger et al. (2018) give is that how-possibly explanations play a central legitimate role in research, in particular, they have three functions:

1. A modal function: how possibly- explanations have value if the phenomenon in question is a modal phenomenon. A modal phenomenon is such that the phenomenon is possibly or necessarily the case cf. Schelling's model of racial segregation. So, Shelling's model is about whether it is possible to understand racial segregation without explicit racial attitudes of the agent. His model shows that it is possible for segregation to emerge with only a minimal racial rule i.e. 30% behavioural rule that says that agents prefer living in non-segregated cities.

2. A heuristic function: the how-possibly understanding that is gained by autonomous toy models is valuable in the construction of other less idealised often more accurate models of the target system.

3. A Pedagogical function: So, some toy models facilitate understanding by providing how - possibly explanation that are grasped by researchers and students (paraphrased from Reutlinger et al. 2018 p.1094).

The claim is that how-possibly explanation plays a legitimate role in research because such explanations have: a modal function that relates the explanation to the phenomenon in question, a heuristic function which is valuable in the construction of less idealised models of the target, and a

pedagogical function which facilitates the understanding of researchers and students by grasping that the relatively simple/ idealised [toy] models provide how-possibly explanation.

4.3.2. Bokulich's 'how-possibly' explanations

Another concept of how-possibly explanation comes from Bokulich (2014). As my concern is with how-possibly explanations, I will limit myself to outlining how how-possibly explanations are generated, returning to her example of the Tiger Bush. I will then show how Bokulich (2014) extends the Reutlinger et al. (2018) conception of how-possibly explanation.

To repeat, the Tiger Bush is a branding of vegetation which has its explanation in computer simulations that are highly idealised. We have seen that the model is argued by Bokulich to provide a how-actually explanation. The question is: how does such a model also provide how-possibly explanation? The how-actually explanation focused on self-organisation raises the further question of what is responsible for the self-organisational mechanism operating in nature, and there is more than one possibility: it could be a deterministic mechanism or a stochastic mechanism of noise-induced pattern formation (paraphrased from Bokulich 2014 p.327). One can now see a branching tree-like structure forming in terms of the how-possibly and how-actually explanations. One of the how-possibly explanations turns into a how-actually explanation and the tree further branches under the appropriate how actually explanation. The further down the tree we go with greater specificity, more work is required of the scientific community to establish the how-possibly explanation as a how-actually explanation. This is because the more general the explanation is, the easier it is to establish it as the one operating in nature.

It should be noted, again, that each level of branching in the tree-like structure of explanations - each level of abstraction that frames the explanandum phenomenon with its corresponding explanatory context-- is done by considering the relevant contrast class of possible explanation. Each level has its particular type of evidence, so that the kind of evidence distinguishing, for instance, between the self-organisational model explanation and the external forcing model explanation is different from that for the next level of specificity on the self-organising branch. Each kind or level of evidence is different to the next level kind of evidence, the evidence that distinguishes between the various specifications of what is responsible for, for instance, the self-organisation model once it has been ascertained as the model operating in nature i.e. a how-actually explanatory model. This elaborates on the nature of contrast classes of how-possibly explanation i.e. they are context bound.

Regarding the how possibly/how actually distinction itself, it is important to recap and note that the following holds for both how-actually (see before) as well as how-possibly explanations: that is, on Bokulich's conception, the distinction itself does not track how detailed an explanation is. So, with respect to how possibly explanations, for Bokulich (2014), it is not the case that they are more abstract than how actually explanations. It is not therefore the case that the more coarse-grained the mechanism is specified, the closer it is to a how-possibly explanation. In fact, the opposite is true: the more finely grained the explanatory mechanism is, the less confident that scientists are that that is the one operating in nature. She extends Reutlinger et al. (2018) in the following way: she shows how-actually and how-possibly explanations interact so a how-possibly explanation becomes a how-actually one. Reutlinger et al. (2018), on the other hand, posits two different kinds of mechanisms for explanation.

4.3.3. The case for how-possibly explanation relating to an impossibility or a how-not-possibly hypothesis

In this subsection, I will also answer the question of how simple models facilitate understanding by generating how-possibly explanations, but in terms of the following refinement: credible simple models present us with relevant possibilities that affect the plausibility of an impossibility hypothesis, enabling us to learn from these simple models (Grune-Yanoff 2009 and see chapter 2). Credible models that show a how-possibly explanation, that is, relevant possibilities that these models express, may affect confidence in a necessity or an impossibility hypothesis.

A credible model gives a relevant possibility in terms of a possible situation which may be inconsistent with an impossibility hypothesis. This happens when a person considers the model and believes in the impossibility hypothesis as well, so for the person to have a consistent belief set, something must change. If the case is that the model cannot be dismissed as irrelevant i.e. the model presents a relevant possibility, then the confidence in the impossibility hypothesis is lowered to accommodate the model. This counts as an instance of learning from the model about the world (Grune-Yanoff 2009). Grune- Yanoff (2009 p.87), to illustrate this procedure of the interaction between the how-possibly and how-actually explanation or the learning effect, gives the example of Schelling's model of racial segregation: Before the Schelling model was published, people believed that racial-segregation was an outcome of explicitly racial preferences. Schelling's model showed that, plausibly, this may not be the case. In publishing the model and its findings, he made people lower their confidence in the racism impossibility hypothesis e.g. that major segregation is impossible without racism.

Firstly, Grune-Yanoff 2009's simple models are credible models just in case they have the power of inspiring belief. To cash out this notion, Grune- Yanoff (2009 p.93) stipulates an analogy between the credibility of models and the credibility of fictions. In particular, the focus is on the assessment of scientific models and the assessment of literary fictions. Fictions express fictional propositions normally true in that work of fiction. And conversely, true propositions may be made false in fiction. Therefore, fictional propositions are "supposed" as true in the work of fiction. Here there are two distinct notions of truth and truth in fiction at play. The definition of fiction is such that it depicts an imaginary situation which is imagined and not believed to be true. The work of fiction has "plausibility" as the criterion for belief. Imagination, moreover, though it has as its origins in fictional descriptions, goes beyond it, so a fictional world is created by filling in the gaps of the description, i.e. in the work of the fiction, by creating a continuous, coherent, whole world of fiction. So, given the degree of sufficiency and coherence in the description, the imaginer will then judge the description to be plausible or credible. In a nutshell, models have two aspects in common with fiction concerning credibility:

1. They are assessed in terms of plausibility or credibility.

2. Judging fiction to be credible is the same thing that scientists do with a model: they imagine a world that the model describes. They then manipulate that situation or world, investigating its internal coherence and coherence with our intuitions. The intuitions in question do not exist independently of the imagined world. For instance, we do not have independent beliefs about the Schelling model of agents' behaviour i.e. beliefs about the behaviour of agents that Schelling assigns to them outside his model. Instead, vague intuitions are played out by considering these imaginary worlds i.e., in the specifics of the

> chequerboard model such as: that that I judge something could have been this way i.e., it is a possibility (closely paraphrased from Grune-Yanoff 2009 p.94).

Furthermore, it is important to note that the object of credibility judgement in fiction is how the fictional environment and the fictional characters in their development fit together. In a similar vein, modellers argue about the credibility of models, so the credibility judgements of the model i.e. judging a model to be credible, establishes itself in terms of depicting a possible world, a scenario how the world could be (Grune-Yanoff 2009 p.95) and they are conditional, for example, an agent's modelled perceptions are credible conditioned on the model environment.

Secondly, at this juncture, the question is: what do we mean by the relevant possibilities of considering certain models affecting confidence in the impossibility hypothesis? This is just the question of how the how-possibly explanation interacts with a how-actually explanation thereby constituting an instance of learning. Knowledge containing impossibility or necessity hypotheses are often established in context of ignorance about which factors are at work, so they are not established by stringent scientific means i.e. they are established in contexts of ignorance. In such contexts people maintain general principles about necessary connection or the impossible coexistence of certain factors that are obvious or intuitive (paraphrased from Grune-Yanoff 2009 p.95). Folk economics is the domain of economics that refers to the set of such principles. These principles have as their origin, "the untrained human perception of economic phenomena" (Grune-Yanoff 2009 p.96). Further, such principles may contradict economic theory and examples of such principles are impossibility hypotheses which are an important part of economic knowledge. An example of folk economics can also be given by Schelling's model of racial segregation insofar as the impossibility hypothesis that major segregation is impossible without racism may be an instance of a principle of folk economics. It is a principle not established by stringent scientific means but one that people find particularly intuitive.

So, Given strong folk notions, as is in such cases, people have a high level of confidence in impossibility or necessity hypotheses.[11] And consideration of a model may have the effect of lowering confidence in such hypotheses: once arriving at the conclusion that the model world is credible, so we say that the model presents a possible situation or world, at least we have this belief. Such a possible model world may exhibit instances that contradict the impossibility hypotheses. But the person who believes in the possible model world also believes in the impossibility hypothesis so something must give, we can either abandon the credible model world possibility or lower confidence in the impossibility hypothesis. But if the model cannot be rejected and expresses a relevant possibility, we are forced to lower our confidence in the impossibility hypothesis.

Now we are in a position to articulate what it means for a how-possibly explanation to affect the how-actually one or what it is to learn from the model. For Grune–Yanoff (2009) the following two conditions must hold:

1. The model presents a relevant possibility, for instance, it is credible.

2. The model contradicts an impossibility hypothesis that as a whole is held with sufficiently high confidence by the potential learners, (closely paraphrased from Grune-Yanoff p.97).

We need 1 plus 2 because 1 without 2 implies we can learn from every credible model. But an arbitrary credible model merely shows the possibility of a state not believed to be impossible in the first instance, which is epistemically futile. However, if conditions 1 and 2 are both held together, credible models help in terms of pitching beliefs about possibilities against beliefs about impossibilities (Grune-Yanoff 2009). For instance, we can generate beliefs about individual credible

[11] it is important to note that impossibility and necessity hypotheses have the same logical form: for all x , if Px then Qx so it cannot be the case that both Px and not Qx.

behaviour on the one hand and on the other hand pitch it against beliefs about aggregate social entities. So learning is by the epistemic surrogates[12] of individual beliefs -- in the model-- and deriving aggregate consequences from them and this derivation thus constitutes an instance of learning.

Another aspect of learning from models for Grune-Yanoff is that world- linking properties are not necessary for learning from models. Grune-Yanoff (2009 p.97) argues that a closer inspection of modelling practice gives the general observation that some economic models do not argue for a model-world link either in constructing the model or establishing it once the model has been constructed. He further stipulates that in constructing such an economic model, economic theorists stress the role of creativity, playfulness, and imagination. He illustrates this with the forming of Schelling's chequerboard model of racial segregation: Schelling wanted to learn about associations or spatial patterns reflecting preferences about whom to associate with in neighbourhoods: "I found nothing I could use as an illustrative material and decided to work something out for myself" (Schelling 2006 quoted by Grune-Yanoff 2009 p.87).

The above quote shows that Schelling could not find any sociological descriptions of the phenomenon he was interested in, so he made up his own surrogate system. He constructed an epistemic surrogate of the situation he was interested in after failing to find the relevant sociological descriptions of the phenomena in in the real world. Note that Schelling also got dynamics of greater interest by modifying the model, so for instance, moving coins in a single line, then in two dimensions and finally settling for coins on the chequerboard. In this way he could get more interesting dynamics (that is, with coins on the chequerboard rather than with coins in a single line). So, in the example of the development of the Schelling chequerboard model, he is not concerned

[12] The epistemic targets are models such that models focus on them in place of the eventual object of interest. One note about the eventual object of interest by constructing manipulating and analysing the models. So, one learns about the eventual object of interest by such surrogates. The reason that Grune-Yanoff labels his simple models as epistemic surrogates is because the models act as surrogates of rea world systems in this way.

with, at least according to Grune-Yanoff (2009), a stringent model world–link and so such the model is characterised as an epistemic surrogate (see chapter 2 and above).

The way in which Grune-Yanoff (2009) extends the concept of how-possibly explanations is in the same vein as Bokulich (2014) i.e., by showing how how-possibly explanations and how-actually explanations interact. However, Bokulich (2014) characterises the two kinds of explanations on a continuum i.e., how-possibly explanations turning into how-actually ones. Grune-Yanoff (2009) on the other hand shows how a how-possibly explanation affects a how-actually one in terms of lowering the confidence of an impossibility hypothesis.

4.4. Explanation and understanding

Reutlinger et al. (2018), are interested in the epistemic goal of toy modelling. Their claim is that one such goal is furnishing individuals with understanding. They help themselves to a theory of understanding that best suits their conception of toy models. This theory is one that allows for understanding from idealised models. The theory is called 'the refined simple view' and is a modification of Strevens's (2013) 'simple view of understanding'. According to the simple view, "an individual has scientific understanding of the phenomena just in case they grasp a correct explanation of that phenomena" (Strevens 2013 quoted by Reutlinger et al. 2018 p.1084). They use Strevens's view because typical accounts of scientific understanding do not meet the challenge from idealised models while his does. That is, typical theories say very little about how understanding from idealised models is possible. This is important because they want a theory of understanding that can accommodate a conception of toy models as very simple and highly idealised models. Such toy models do not meet the veridicality condition for typical theories of understanding. The

veridicality condition states that the explanatory assumptions, i.e. the explanans, needs to be true or approximately true using Strevens's 2013 account.

The Reutlinger et al. (2018) theory of understanding refines Strevens (2013) simple view of understanding: How-actually explanations can yield understanding in a variety of ways using any one of the received views of idealisation and their justification. Most important, however, is that the interpretation and justification of idealisations in the model meet the veridicality condition, and this may be done by applying a reinterpretation of them. For instance, in Strevens's case, idealisations prior to reinterpretation are simply false of their target. A reinterpretation for instance using McMullin's view (see Galilean idealisations in chapter 2) would make them approximately true of the target. Strevens (2008) uses a "minimalist" interpretation and, according to Reutlinger et al. (2018), he says that the idealisations in the model do not matter. In this sense the idealised assumptions refer to explanatory irrelevant factors, so, this minimal idealisation account meets veridicality in so far as the idealised models truthfully represent the fact that some factors (idealisations) are not explanatorily relevant.

The most important refinement for my purpose is that there are different modalities of explanation, and therefore of understanding, because there are at least two types of explanatory information gained from toy models---- that which allows for how-actually explanation and that which allows for how-possibly explanation. How-actually explanation satisfies the veridicality condition whereas how-possibly explanation identifies only possible explanatory factors.

We can now give the core statement of the refined simple view:

Scientist S understands phenomenon P via model M in context C if one of the following conditions hold:

1. Scientist S has how-actually understanding of phenomenon P via model M in context C if and only if model M provides a how actually-explanation of P and S grasps M.

2. Scientist S has how-possibly understanding of phenomenon P via model M in context C if and only if model M provides a how-possibly explanation of P and S grasps M.

(Reutlinger et al. 2018 p.1086).

This way of thinking about the understanding/explanation link can be employed for all the views of how-actually and how-possibly explanation with toy models discussed above. Grune-Yanoff and Bokulich do not focus on a theory of understanding, but insofar as they identify how toy models can be used to obtain explanations, we can also link their analyses to claims about understanding if we adopt the Reutlinger et al. refined simple view.

With respect to grasp, as noted in chapter 3, Reutlinger et al. (2018), interpret grasp as the fundamental relation between mind and the world; it is a philosophically fundamental relation (philosophically primitive) though not scientifically primitive. This subjective component can thus be further investigated by cognitive science and, in effect, Reutlinger et al., naturalises Strevens's (2013) notion of grasp. Strevens himself comments (in conversation and in a draft paper in preparation Strevens 2020), that for true representation i.e., for models that represent the target, the grasp story he advocates straightforwardly applies (see chapter 3). So that grasping the model is grasping the real state of affairs asserted by the model to exist, and thereby also grasping the connections between them. Grasping for how-possibly explanations would work in Strevens's fictional case (Strevens draft 2020) where grasping is equated with the grasp of unreal things in the world of the sentences of the how-possibly explanation that would have been true, and then test for grasp there, (for greater elaboration, see the section on grasp in chapter 3).

4.5. Conclusions

In this chapter, I explored one prominent answer to the question of how simple models can facilitate understanding: by providing how-actually and how-possibly explanation that researchers can grasp. Different people expand on the how-possibly and how-actually distinction in different ways, though Reutlinger et al. (2018) provide the overarching conception, positing two distinct types of mechanisms: the first one is a partial how-actually explanation (how possibly-explanation) and the other is a how-actually explanation. Both Bokulich (2014) and Grune-Yanoff (2009) have distinct ways in which how-possibly and how-actually explanations interact with each other, thereby extending the Reutlinger et al. (2018) conception.

Furthermore, I believe that one cannot draw a sharp line between how-actually and how-possibly explanations because of the variety of competing views that Bokulich (2014), Grune-Yanoff (2009), and Reutlinger et al. (2018) provide suggests that agreement on how to distinguish how-possibly and how actually explanations may be difficult.

5. Simple models facilitate understanding-- the case for minimal model explanation.

5.1. introduction

In chapter 2, I distinguished toy models from Batterman's (2002, 2009, Batterman and Rice 2014) minimal models. In this chapter I will explore one way that his minimal models can facilitate understanding, namely, by providing minimal model explanations. Though understanding on this view is still achieved via explanation (as in the last chapter), I will argue that is a different kind of explanation (and understanding) than we saw in the case of toy models, contrary to what is suggested by Reutlinger et al. (2018). Further, I shall argue that minimal model explanation provides a species of understanding as unification, but where unification is achieved differently than in Philip Kitcher's influential account. This understanding is roughly a localised unification account and will be expanded on in the final section.

My focus will be minimal models as discussed by Batterman (2002, 2009, Batterman and Rice 2014, Batterman 2018). To recount what a minimal model is, it is a type of idealised simple model whose structure explains, in part, patterns of macroscopic behaviour that are heterogeneous at smaller scales, e.g., biological patterns that range over heterogeneous populations. This means that the details do not matter for a proper characterisation of that macroscopic pattern of interest.

Employed in physics as well, a minimal model is one that most "economically caricatures the essential physics" (Batterman 2009 p.430) in cases like this. In what follows I will first discuss Batterman's view of minimal model explanation (section 5.2) and then turn to minimal model understanding (section5.3).

5.2. Minimal model explanation

In this section, after elaborating on the core concepts in Batterman's view, I will then present examples from different fields to show that minimal model explanation is not confined solely to physics (Batterman's main examples) and can be extrapolated to, for instance, the field of biology. I will then highlight the differences between explanations grounded in common features, which can arguably be said to subsume toy model explanation (Reutlinger et al. 2018), and the minimal model account of explanation, which Batterman and Rice (2014) argue is superior to the common features accounts in the sorts of cases discussed.

5.2.1. Basics: phenomena, models, methods

In this section, I will give an overview of Batterman's view in general terms. A Batterman-style minimal model is meant to be definitive of minimal models as I use the term in this thesis. A minimal model is in the class of explanatory models whose explanatory structure has been thus far often misunderstood (Batterman and Rice 2014). Rather than representation, another story about how the models latch on to reality needs to be told. Such a minimal model can be used to perform a

derivation that a repeatable phenomenon, a pattern, like the sex ratio or droplet shape, applies in an arbitrary system that has specific features highlighted in the model. For instance, for the shape of breaking drops, the relevant features are similarity solutions to the 1- dimensional Navier-Stokes equations or in the case of a biological model explaining the sex ratio, the common feature is the substitution cost.

The kinds of phenomena that are of interest, more specifically to be called the 'phenomenon of interest', lend themselves, as Batterman (2009 p.429) contends, more readily to the non-traditional view of the role of idealisation in explanations involving mathematical modelling. The non-traditional view of the role of idealisation argues that idealisation plays a necessary and ineliminable role in the proper explanation of the phenomenon of interest.

The phenomena of interest that Batterman addresses stem from the fact that nature presents us with patterns – repeatable phenomena. We thereby try to understand how those patterns come about -it seems that one of our interests is in repeatability as a salient feature and we thus ask what is responsible for that salient feature?

A goal of mathematical modelling is often to capture this very salient feature of particular phenomena – their repeatability – in a formula. Therefore, the repeatability of the phenomenon of interest is a feature and a constraint on the model. Idealising is a way to focus on that repeatability – those features that are repeated at different times and places (like the shape of a droplet dripping which is repeated no matter whether from a tap or sprayed by a wave) that make up the regularity. The process of idealisation, in virtue of focusing on those patterns thus understood, is a means of removing irrelevant details – those details that are distracting from this focus. More specifically, the details are such that they can change but do not affect the repeatable behaviour of interest. In this way we get a better understanding of what we are interested in, i.e., the repeatable behaviour. Such patterns or repeatable phenomena include the shape of breaking drops, criticality, and the one-to-one sex ratio in biological populations.

It is important to note the difference between abstraction or relatedly "black boxing" (black boxing can sometimes be seen as a kind of abstraction) and the process of idealisation in order to arrive at a representative equation. One uses abstraction techniques to find a representative equation and then one can perform the asymptotic procedure. The abstraction in the sex ratio for example is done in arriving at the representative equation for the 1:1 sex ratio. The representative equation 'black boxes' the details of the physics. For example, in the example of the universality of critical point behaviour, the model captures the diverse behaviour of each of the different systems by a small number of parameters, thus black boxing these behaviours. As understood in this discussion, black boxing means representing the overall effect but hiding/ignoring the ways it is achieved. [13]All this is done before one uses asymptotic techniques to show that such a model, regardless of the values of the parameters will behave the same way, for instance, near the critical point.

By contrast*, idealisation,* as I use the term, is done in these kinds of models at the point where asymptotic techniques are used in delineating a universality class. It is the process of removing irrelevant details. In the sex ratio example, it is "why" we get the substitution cost as Fisher's (original) explanation for the 1:1 sex ratio. The difference I mean to highlight in these cases is roughly this: when you abstract (as I use the term), you come up with a new description that represent a lot of different things under the same heading. This description is not true of any of them, but it implies that the relevant essential facts that are true. We abstract when we use the label 'forest' rather than describing all the trees. In the models I am talking about here that happens at the earlier stages. In idealising you ignore facts. In this case you are idealised in arriving at the few features they all have in common as we do in these models later in finding the universality class.

[13]Though black boxing may be the right idea in this particular case, the idea is often broader than that when used by others including Strevens.

Some people (notably Michael Strevens) have a different view: they use the term idealising to refer to what happens in the earlier stages of model building where representations are produced that are not literally true to the facts in which a lot of black boxing and idealising takes place. Idealising, as I use the term is distinguished from black boxing in the following way: scientists first black box, then idealise by saying that the black boxes are behaving the same way.

The view of Strevens and others, states that in model building there is a lot of black boxing and idealising, for example the differences between heterogeneous systems is captured by a small number of parameters. Then, on this view, the asymptotic techniques show that any such idealised model, regardless of the values of its parameters, will behave in the same way near the critical point. This then is the explanation of the universality of critical point behaviour, for example, in the case of the Ising model. The asymptotic techniques then, do not represent idealisation in the usual sense or as I use the term.

In (Batterman's) minimal models, the explanans is supposed to be a mathematical derivation using idealisation of the phenomenon of interest. From inspection of the derivation, it is apparent that different micro-states are all the same at the macroscopic level. Their details are irrelevant and do not get in the way of a model's derivation of the phenomena of interest. Such details may even impede a proper understanding of the phenomenon. This view, 'the non- traditional view', of mathematical modelling, also points out that adding detail will not improve the minimal model.

In Batterman-type cases, the feature of the model that allows derivation of the same macro-state, across different micro-states, involves the mathematical operation of taking limits. The model offers a representative equation. This mathematical operation of taking limits removes irrelevant details and constitutes the process of idealisation, given the representative equation. There is a

recipe[14] for getting insight from the model's representative equation,[15]of which taking limits is a part. This recipe states a set of procedures that best characterises the modeller's method of simplification (paraphrased from Batterman 2009 pp.430-431).

Two important procedures are firstly that the modeller non-dimensionalises[16] the representative equation allowing her to compare the parameters appearing in the equation to judge their "size", even if they were once expressed in different units. Secondly and most importantly, the modeller takes limits of a given parameter thus reducing the equation. The taking of limits is constituted by letting a small, non-dimensionalised parameter approach zero and for a large non-dimensionalised parameter to approach infinity. Infinity and zero are both limiting values and one simplifies by idealising in this way. This is an analytical procedure that plays an important part in the scientific investigation of certain types of phenomena, a scientific investigation that does not care for accurate numerical predictions/computation. Rather what we are after is a minimal model, a model which displays the dominant features of the system. Such a limiting model shows the essential physics or whatever the relevant science happens to be. As we will see, examples of these models include the minimal model of breaking drops, the minimal model of the one-to-one sex ratio, and the minimal model of criticality.

In each case, the derivation and analysis using the minimal model explains:

[14] Such a recipe is by and large independent of any particular view such as the non- traditional view of modelling though Batterman thinks that this set of processes fit the non-traditional view best.
[15] Insight is gained by delimiting a universality class---a demonstration that details that distinguish the model system and different real systems are irrelevant.
[16] It should be noted that strictly speaking non-dimensionaling say the equation is doing more abstraction than is needed for Batterman's purposes see figure 5.1 where the reduced dimensional coordinates means abstracting from scale but not for instance temperature. However, this would happen if we de-dimensionalise the quantity as physicists would do.

1. The phenomenon in question itself- the large-scale patterns or macro-behaviour e.g. the sex ratio or droplet shape.

2. Why the phenomenon in question occur in a system with the highlighted features e.g., in an arbitrary population with specified features in the case of the sex ratio example.

In the minimal model, these highlighted features, such as the linear substitution cost and the solutions to the 1-dimensional Navier Stokes equations, are necessary in the derivation and thus the model also explains:

3. Why these features are common to systems exhibiting these phenomena.

The approximations employed in the models are to coarse grain across other features making them inessential for deriving these phenomena and thus also explain:

4. Why variations in other features are irrelevant (loosely paraphrased from Batterman and Rice 2015 p.270).

5.2.2. A paradigmatic example

This sub-section explores a paradigmatic example of minimal modelling, that is the minimal model of criticality. Universality is the feature by which a minimal model explains and is a species of asymptotic methods in general. The essence of universality can be summed up as the fact that many systems, despite lower level heterogeneity, exhibit similar or identical behaviour at a higher level. More specifically, there are two general features that characterise universality:

1. The details of the system that would feature in a complete and causal mechanical explanation[17] of the system's behaviour are largely irrelevant for describing the behaviour of interest.

2. Many different systems with completely different micro-details will exhibit identical macro behaviour (closely paraphrased from Batterman 2018 p.864-5).

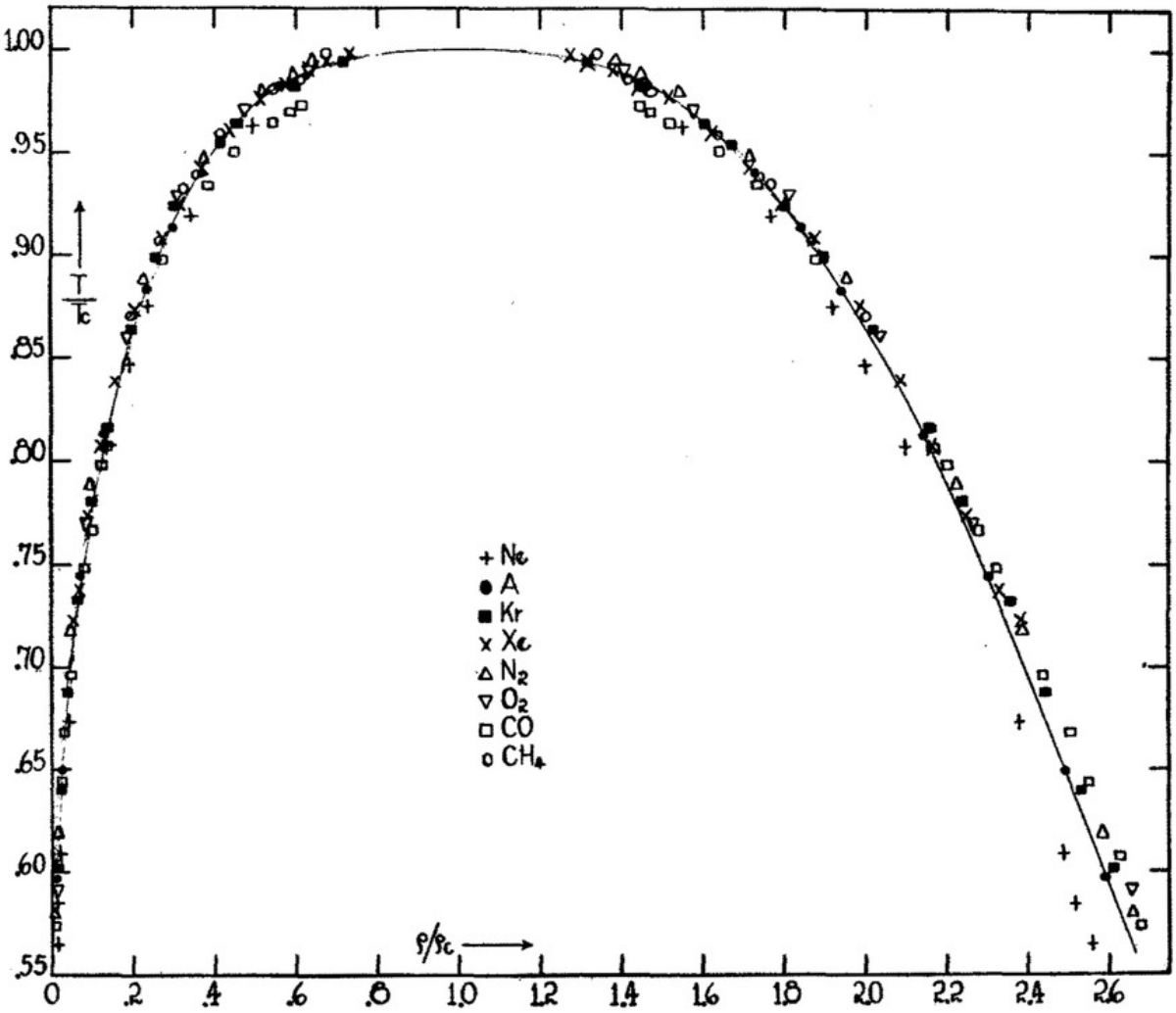
Again, it is important to distinguish the abstraction that is done when constructing a model from the asymptotic operations performed on the model, in order to, in this case, explain critical point behaviour. In constructing the model, one uses a lot of abstraction, so that the differences between the systems are represented by a handful of parameters, and only then the asymptotic techniques can be applied.

Criticality, as noted above, serves as a helpful example to illustrate universality near critical points. Consider a curve on a graph that plots temperature versus density from eight different fluids in reduced dimensional co-ordinates such that the values on the X axis below a certain point, call it P, represent the density of vapour phase of fluids and, above P, represent the density of the liquid

[17] The causal mechanical approach will be expanded on later in contrast to Batterman's conception of minimal model explanation.

phase of fluids. At point P, the densities of different phases are the same. The y-axis plots a dimensionless critical ratio of temperature to critical temperature of the fluids. The curve on the graph provides the different densities of liquid and vapour phases at different temperatures. The shape of the curve is the same for each fluid (each with different molecular composition). The pattern depicted in the graph is thus a paradigm example of universality (paraphrased from Batterman 2018 p.864).

Figure 5.1. Temperature versus density from 8 different fluids in reduced dimensional coordinates to show the universality of critical phenomena (Batterman 2018, Guggenheim 1945).



Asymptotic methods in general, provide principled reasons grounded in the fundamental physics of the system as to why many of the details that distinguish systems from one another are irrelevant when it comes to explaining the universal behaviour of interest (the upper level generalisation is the universal behaviour). Batterman (2018 pp.864-865) outlines his explanatory strategy with respect to the thermodynamic behaviour of systems near criticality: the basic idea is that universal upper scale thermodynamic behaviour of systems near criticality is dominated by fluctuations in various crucial quantities, and the range of those fluctuations is much greater than the range of intermolecular forces governing the interactions between molecules. Therefore, the fluctuations are insensitive to and dominate the detailed nature of intermolecular forces (essentially characterising the fluid).

This can be spelt out further with a renormalisation group argument showing that in certain limits, the differences in intermolecular behaviour play essentially no role or do not affect the upper scale behaviour. This renormalisation methodology involves morphing one system into another. “Morphing” may be done in the following way: first one constructs an abstract space, each point representing for instance, a real fluid, a possible fluid, solid et cetera. Secondly, one introduces a transformation on the space that has the effect, upon iteration, of reducing the degrees of freedom by some kind of averaging rule, replacing, for instance, specific interactions among a cluster of nearby molecules with averages. Thirdly, one considers these averages to be the new molecular components that interact with other molecular components in a new coarse-grained cluster. In effect the averaging rule yields a new coarser grained system.

It turns out that, near the critical point, systems exhibit the property of self-similarity, which guarantees that the new system looks like the old one as one coarse grains by replacing the original degree of freedom with averages. Furthermore, by examining the topology of this renormalisation group transformation (coarse graining) on the abstract space, you will eventually find the fixed points of the transformation. A fixed point is a property of the transformation such that the details of the systems that flow to that fixed point have been removed. Universality classes are then therefore delimited or specified given a procedure of specification by systems that flow to a fixed point. Such systems, therefore, that are in the same universality class, display the same macro - behaviour and so the transformation, e.g. by renormalisation group methods, in the neighbourhood of the fixed-point, shows what the macro-behaviour must be. Thus, this is a recipe for how to build explanatory models that show how universality arises from the lower level. It is a paradigmatic example of the use of asymptotic methods.

Section 5.2.3. Further examples

Breaking drops.

This example focuses on the breaking of droplets from the point of view of analytic fluid dynamics. The analysis generates an explanation whereby the details drop out. The repeatable behaviour of interest here is the universal shape of a breaking drop. The idealising procedures of the minimal model show that breaking drops, whether the drop drips from a tap or is sprayed in the air by a crashing wave, are members of the same universality class. Batterman's analysis shows that the various fluids' micro-constituents fall out of the picture for characterising the shape of breaking drops; the evolution of the fluid before and after breakup is independent of the molecular micro-constituents. The description below follows that of Batterman 2009 (pp.435-438).

Again, it is important to notice the abstraction involved before the asymptotic procedure takes place-- in this case, it would be the abstractions involved in analytic fluid dynamics-- and to contrast this with what I dub as the process of idealisation. Idealisation, as I use it, occurs after one has arrived at a representative equation(s), that has already been treated by black-boxing or idealising (in the other sense) techniques.

To characterise the shape of the fluid interface at and near breakup, of say a tap dripping, one needs to examine the Navier-Stokes (NS) equations for free surface flows. These equations, at the point of breakup, develop a singularity in finite time. The singularity signifies a breakdown of these equations and breakup is characterised by divergences in fluid velocity and the curvature of the interface at the point of breakup. The NS equations themselves have two boundary conditions, implying that the surface moves with the fluid at the boundary. These equations define a complex moving boundary value problem given the fact that we are interested in what happens at fluid breakup, and it seems that this problem is made much more difficult because non-linear effects will dominate (paraphrased from Batterman 2009 pp.434-5).

However, if one focuses on what happens at singularity, it is possible for the problem to be simplified and to provide an exact solution to the moving boundary problem. There are two features of this problem that allow this to happen: firstly, there is the assumption that the singularity is line-like enabling one to find a one-dimensional solution to the full NS equations. Further, we can non-dimensionalize the quantities appearing in the problem by introducing a characteristic timescale (which however will be different at various stages of the singularity). The second feature of the moving boundary problem is that near singularity, surface tension, viscous forces and inertial forces all become equally important, making the fluid acceleration diverge and in effect leaving the constant acceleration of gravity out of the picture.

Moreover, close to singularity, and without going into technical details, what is crucial, is that the singular solutions of the one-dimensional NS equations have similarity properties (see below). For instance, if one introduces the viscous length and timescale, such scales imply that the viscosity is doubled, and breakup will look the same for length scales four times as large and for timescales eight times as large. This is a similarity property and the existence of such a similarity solution indicates the robustness and stability of the phenomenon under perturbation of details.

The existence of a similarity solution (in the similarity variable) indicates the universality of the shape function of breaking drops. The similarity solution characterises the universality class of fluids at breakup. The expectedness of these shapes is given by the nature of the similarity solution. Moreover, for larger scales, the similarity solution continues uniquely from before to after breakup, so the conclusion, to reiterate, is that the evolution of the fluid before and after breakup is independent of molecular microscopic details e.g., the different fluids with their differing viscosities dripping from different nozzles etc. (closely paraphrased from Batterman 2009 p.438).

Fisher's sex ratio model.

This example shows that minimal model explanation is not solely relegated to the domain of physics but can occur in other sciences too. One example from biology uses a simple [minimal] mathematical model----Fisher's sex ratio model –----- reinterpreted to explain the behaviour of a diverse range of populations without referring to the details of the populations. Batterman and Rice (2014) contend that Fisher's original model has some important features that have either been ignored or misunderstood and that a proper understanding of Fisher's sex ratio model is as a minimal model. In order to see this, we need to first present Fisher's original conception which is roughly that if a population moves away from a one-to-one sex ratio, there will be a fitness advantage favouring the parents who overproduce the minority (paraphrased from Batterman and Rice 2014 p.367). That is, there will be a fitness advantage until the sex ratio re-balances to a one-to-one ratio, i.e. a stable equilibrium of the evolving population.

Fisher's original model relies heavily on a key trade-off known as the substitution cost between the ability to produce males and females. The trade-off, which is Fisher's explanation of the one-to-one sex ratio, is linear, meaning that males and females cost the same resources to produce, so one fewer male means one more daughter. The commonality of the linear substitution cost explains firstly the one-to-one ratio across many natural populations and the behaviour of other systems near the fixed point of the one-to-one ratio. This is the "explanation" (for Fisher) and his original model that I have been referring to throughout and should be distinguished from the asymptotic explanation that is the minimal model.

However, Batterman and Rice (2014) contend that just citing this common feature as a relevant feature is not enough for a proper explanation of the phenomenon in question: the one-to-one sex ratio. They contend that there are several additional assumptions that may be at play and that really the explanation provided by Fisher's model involves far more and far less than the highlighted

features, like the substitution cost, that it has in common with real populations (Batterman and Rice 2014 p.368). It involves additional idealising assumptions such as infinite populations and asexual mating and is therefore only a caricature of real populations. The function of such a caricature is to gain insight into key variables i.e. the common features. Furthermore, the additional assumptions show extreme misrepresentation when it comes to the behaviour of any real biological populations, but the model remains explanatory; the question is, how can this be the case?

The answer: Fisher's model is a minimal model. A minimal model is a caricature and, as such, an essential ingredient in its explanation lies in the method for delimiting the universality class for the 1:1 sex ratio for real biological populations and showing that Fisher's model population is in that class. This tells us "why" those features, common features across populations, [i.e. the linear substitution cost] are common and relevant; just by citing the common features as common features is not enough to answer the "why" question at hand. Unpacked, such a "why" question is comprised of 3 sub questions: *why* are the common features necessary for the phenomenon to occur? *Why* are the remaining heterogenous details irrelevant for the occurrence of the phenomenon? and *why* do very different populations have these features in common? These correspond to items 2, 3, and 4 of the explananda types described at the start.

To demonstrate this, Batterman and Rice (2014 p.372) start with a space of possible systems, where some points represent real populations and other points represent model populations. What distinguishes between the populations is the type of organisms, assumptions about evolutionary dynamics, population size (infinite or finite) etc. In order to delimit a universality class we can use a variety of techniques more or less analogous to the renormalisation group methods (see above) to show that the model population is in the same universality class as the real populations whose sex ratio the model tries to explain.

The minimal model provides a demonstration that shows why it is that the common dominant feature, the substitution cost, is a common feature. In order to do so, we need a story as to why the

details do not matter but these features do. In doing so we understand why these features, the common features are necessary for the phenomenon to occur. Given that we can delimit this class of systems that exhibit the one-to-one sex ratio from other different systems, delimiting the universality class accounts for the safety of using a minimal model to explain the relevant patterns over heterogeneous biological populations. This demonstrates the stability of the sex ratio given changes in details of the systems involved. So, we can perturb the details e.g. by changing a population of sheep into a population of deer, without affecting the stable sex ratio these populations display.

Given, also, that Fisher's model population is seen to be a member of this class, this means that its details could also morph into that of actual specific populations without a change in the sex ratio. The stability under these kinds of perturbations of this feature answers the question: why are the heterogeneous details of the systems irrelevant to the evolution of the 1:1 sex ratio? Consequently, we also discover that the linear substitution cost feature is common to the members in the same universality class; it is a common feature despite heterogeneous variation between the systems. This also shows us why those features are common features; it shows why various heterogeneous systems display the linear substitution cost. Finally, by demonstrating that all the members in the universality class do indeed have these features, one answers the question: why are these common features necessary for the phenomenon to happen? And as a result, one also has the explanation of why other different systems fail to have these features (fail to display the macro-scale behaviour). In this way Fisher's minimal model is one way in which a minimally accurate model helps us to explain the behaviour of interest in real systems.

So, delimiting the universality class explains *why* different populations have these features (the common feature) in common. Also, by delimiting the universality class we get a story as to *why* the details distinguishing one system from another do not matter. And in doing this we also see *why* it is that only these common features are the relevant ones. Furthermore, delimiting the class of systems

that display the 1:1 sex ratio is also delimiting a universality class that is safe for using minimal minimally accurate models, for example Fisher's 1:1 sex ratio model, as well as showing the robustness or stability of the 1:1 sex ratio under perturbation of details. All this gives us an answer as to *why* Fisher's model (as a minimal model) can be used to explain the behaviour of biological populations that display the 1:1 sex ratio, that is, because the population in the model with the characteristics ascribed to it there, is in the same universality class as other systems that have the 1:1 sex ratio as a stable equilibrium.

5.2.4. Common features versus minimal model explanation

Batterman and Rice (2014) insist on the difference between common features accounts of explanation (encompassing mechanistic and the difference making views – these kinds of explanation can be said to be like that which Reutlinger et al. 2018 subscribe to) and the minimal model account of explanation. Furthermore, they argue that minimal model explanations are often better than those of common features accounts.

The common features account broadly entails a claim to the effect that the model's explanatory force is gained by the model having certain relevant features in common with its target. Therefore, a model, on this view, is explanatory just in case it accurately represents the relevant features of its target (Batterman and Rice 2015 p.351).

Examples of the common features accounts are mechanistic accounts, for example those discussed by Kaplan and Kramer (2011), where there is a correspondence to the actual causal mechanisms responsible for the phenomenon to be explained, as well as causal or difference-

making accounts (e.g. Strevens 2004, 2008). The distinguishing feature of the common features account in general is that it has accuracy requirements. Such requirements mean for instance, in the case of mechanistic models, that the explanatory force of the model consists in satisfying those accuracy requirements e.g. correspondence of the causal mechanism, described in the model, to the actual causal mechanism in the target. In the case of causal difference-making models, there is still an accuracy requirement, only weaker, insofar as an explanatory model should accurately represent the difference-making causes that produce the phenomenon to be explained. Furthermore, the accuracy requirement in the case of mechanistic models means that the more accurate in detail the model is with respect to the items that explain, for instance in corresponding to the target, the better the explanation.

It is also plausible that the how-actually explanations that Reutlinger et al. (2018) discuss are often mechanistic, causal or difference-making in character (see chapter 4). This is because the toy model account uses Strevens's (2013) simple theory of understanding in which their explanation clause is embedded; and according to Strevens the explanation clause involves a veridicality condition. Idealisations are merely reinterpreted with a suitable theory of idealisation to make them approximately true or compatible with the veridicality condition (see chapter 3 for further discussion).

So the main difference between the common features account and the minimal model account is that the important feature of the common features account is the model's common features with the target, while for the minimal model account, it is that the model system falls in the same universality class as other systems that share the common feature. So, the *way* in which the two accounts explain is the key difference.

Batterman and Rice (2014) contend that minimal model explanation is sometimes a better explanation than that of the common feature account, for several reasons:

- The common features account conflates the following: what accuracy conditions are required for the model to explain and in virtue of what does the model explain?

- For the case of mechanistic models, a species of the common feature account, too much weight is given to accuracy requirements. Batterman and Rice (2014) stipulate that many models are explanatory despite not actually describing or representing the real causal mechanisms that produce the phenomenon of interest. Also, with respect to increased detail as model improvement, they contend that for some models, the explanation given by the model is improved by removing details of the causal mechanism.

- In the case of causal difference-making models, also a species of the common features account, not enough positive contribution is made by the idealisations in the model. This means that for instance in Strevens's 2008 account of difference-making models, idealisation plays a meagre role insofar as those idealisations do not provide any real explanatory force. They merely highlight the causally irrelevant factors of the model.

- The common features account fails to answer questions that are essential for good explanation, such as why these common features are necessary for the phenomenon to occur (see above).

- The common features account only cites the features in common between the model and the target system. A minimal model provides vital information that the common features models lack by showing that the minimal model system falls into the same universality class as other systems that contain those common features (paraphrased from Batterman and Rice 2014 p.351).

To illustrate this, concerning minimal models, continuing from the above example of breaking drops (Batterman 2009), Batterman stipulates "why" the minimal model gives us an explanation of the phenomenon of interest. That is, why is it that upper-level theory, e.g. analytic fluid mechanics that contain idealisations in contrast to the more fundamental theory, is explanatorily ineliminable to a proper explanation of breaking drops? Ineliminable in this context means that the more detailed explanation from fundamental theory-- in this case molecular dynamics simulations at the nano level --does not provide a complete explanation alone without a vital contribution from upper-level theory. We have seen how explanation is achieved by analytical fluid mechanics in the example of breaking drops (above, section 5.2.1).

In contrast we also get "an explanation" from the more fundamental theory of molecular dynamics. Batterman (2009 pp.443-446) recounts the experiment of Mosler and Landman (2000) who investigate the breakup of jets at the nano (fundamental) level by simulation. They find that a typical shape at breakup resembles a double cone shape. However, by applying the hydrodynamic equations, the simulated breakup shape develop long necks prior to breakup. So, there is a discrepancy between the double cone shapes of the nano-jet simulations and the shapes with long necks of the continuum hydrodynamic equations. The discrepancy with molecular dynamical description and the hydrodynamic description of the same process shows that continuum deterministic hydrodynamic equations fail at the nano scale. Large fluctuations in the

hydrodynamics become important at the nano scale signalling a breakdown of the deterministic continuum description, so we have singularity.

However, by introducing a stochastic term (Gaussian noise) into the hydrodynamic equations, one can produce a solution to the stochastic continuum equations which agrees with the double cone shape of molecular simulations. The question now is 'given the statistical nature of the double cone shape of the molecular simulation, why is this shape statistically expected?' This "why" question may be answered by the similarity solutions to the NS equations for the hydrodynamic description, therefore one has an answer for an analogous explanatory "why" question on a larger scale: one explains or understands why the given drop shape occurs at breakup and also why it is expected.

Singularity is appealed to in order to answer this question. Without singularity, there is no similarity solution giving the explanation of universal behaviour. Therefore, the breakdown of the continuum equations allows for the explanation of universality. Moreover, no such explanation is available from fundamental theory: in fundamental theory, there is no appeal to singularity to explain the statistically universal double cone shape of breaking drops. From the viewpoint of fundamental theory, there is no breakup, the location of breakup in space or time is absent. Details of the molecular dynamics therefore drop out because the breakdown of the continuation equations enables us to provide explanation of the universal shape -- the double cone shape of breaking drops.

Given the double cone shape expectedness, however, we still need an account of the statistical universality of this shape function in nano-jet breakup at the more fundamental level like the one found at larger scales with the similarity solutions. At the nano level, the solution is generated by random noise introduced into the continuum equations. Most importantly, however, the solution needs to assume for a fixed breakup time that it is self-similar i.e. that there is a similarity solution. As we have seen this is only possible given the assumption of singularity at the fixed breakup time in

the stochastic hydrodynamic equations (in the less fundamental theory of hydrodynamics). In effect this is an explanation for why the double cone shape function is expected in molecular dynamical simulations. To reiterate: this explanation is grounded in the less fundamental theory of hydrodynamics (paraphrased from Batterman 2009 pp.443-446).

5.3. Minimal model understanding

The last subsection explored minimal model explanation and contrasted it with the common features account. In the remainder of the section, I will consider the character or nature of minimal model understanding. My contention is that such understanding has a unificationist character.

Batterman's remarks related to understanding appeal to asymptotic analyses and to the "stability of mathematical structures and appropriate abstract spaces" (Batterman 2002 p.35). I will try to supplement Batterman's remarks to develop a richer characterisation of the understanding gained via minimal models. Such an account of understanding should elaborate on the *why* question as elaborated on throughout this chapter.

To put minimal model understanding in context: I will adapt Strevens's (2013) schema of simple understanding[18]: that understanding involves grasping a correct explanation. The adaptation consists in stipulating explanation as minimal model explanation.

Understanding that X is in a universality class[19] requires the agent to grasp that X is in a certain universality class. In this way grasping a minimal model explanation would also be, on Strevens's

[18] It should be noted that Reutlinger et al. 2018 use a refinement of the same schema for how-possibly and how -actually understanding.
[19] This is a statement of understanding for minimal model understanding using asymptotic analyses.

account, grasping an explanation that is partly or wholly fictional. It is partly or wholly fictional because of the ineliminable aspect of idealisation in Batterman's minimal model explanation, along with the fact that they, the idealisations, have explanatory force. So, the question now is, how do we grasp so-called fictional models? According to Strevens's (2020) account, one way to do this is if we were to go to a world where the sentences of the model would have been true and then test out the thinker's recognitional powers there (see chapter 3 for further elaboration). In this way grasping is something that someone is capable of doing if a certain state of affairs would obtain. I think that something like this is going on in Batterman's minimal model explanation, because the model assumptions are distorted and false, and the idealisations plays an ineliminable role in explanation, therefore the minimal model explanation could be seen as falling into the fictional case of grasping.[20]

The correctness of a minimal model explanation lies in the *way* the explanatory model latches on to the world, e.g., the demonstration of the relevant model population, which exhibits the 1:1 sex ratio, falling into the same universality class of other populations that also exhibit the 1:1 sex ratio (both real populations as well as abstract ones). To illustrate this more clearly, I will provide an example consisting of two different curves explaining two different systems in the same class i.e. showing that the explanation is derived from the same set of axioms. In this case we have two different explanations showing *that* each system is in that class but there lacks an explanation as to *"why"* each of those systems is in that class. Batterman's asymptotic analyses does exactly this. To introduce asymptotic analysis with a brief metaphor, one can say that asymptotic analysis is the mathematical ladder to identifying which universality class one's case is a member of.

So far, we have discussed causal-mechanical explanations as a species of the common features account above. However, it is worth noting that Batterman (2000 pp.230-233, 252- 254) elaborates

[20] This could also work with the simple psychological case of grasping for fictional models but the recognitional account is more recent

on the causal -mechanical approach in contrast to the unificationist view – both, consequently different to the asymptotic analyses he endorses, as introduced above.

The causal-mechanical view has been characterised by Wesley Salmon (1989) as bottom-up and the unification view as top-down. The difference is a difference in their aim for explanation. The causal-mechanical explanation also appeals to the underlying microstructure of the phenomenon of interest. These are local explanations in that a phenomenon is explained by the collection of causal processes that bring it about. The causal mechanical view admits a kind of local understanding. Understanding is achieved when one can elucidate the causal mechanisms operating in an opaque black box. Explanatory knowledge reveals the black box that holds nature's inner workings. Such inner workings can be provided by Peter Railton's ideal explanatory text (Kitcher 1989): which would be the whole story for the explanandum relative to the correct theory of the world. The whole story would need to be a complete detailed description of the causal mechanisms involved as well as the theoretical derivations of all the relevant governing laws. So, the idea is in the ability to generate this ideal text to give understanding of a given phenomenon.

Understanding on the causal-mechanical view requires detail and precision. So understanding is gained on this view only when the detailed mechanisms are exposed. The use of formal arguments is subsidiary to this task. This is a fundamental difference between the causal-mechanical view and the unificationist view of understanding. On the unificationist view, understanding arises when we recognize that diverse phenomena can be derived using the same argument forms, and the diverse phenomena are thereby unified. It is a reduction in the number of the arguments that provide unificationist understanding. So, the role of argument and derivation is more essential to unificationist view of understanding than to the causal mechanical view.

The general sentiment captured by various unification theorists (Friedman 1974, Kitcher 1976) may be summed up by the following statement:

The explanatory status of a particular argument is in terms of its relationship to unifying theoretical structures that reveal connections amongst diverse phenomena (Woody 2013 p.10).

And so unificationist understanding naturally comes from seeing connections in different kinds of phenomena. This captures something essential about explanatory unification. The example above of criticality provides an illustration. Belonging to the same universality class is the connection unifying the different types of fluids. Figure 5.1 demonstrates the universality of critical phenomena showing (in reduced dimensional coordinates), that near critical points, the shape of the curve that plots temperature versus density is the same. This shows that universality explains the upper scale (macro) thermodynamic behaviour of different fluids near criticality.

However, Batterman provides a different conception of scientific understanding to both the causal-mechanical view and the unificationist view. Both these latter views emerge from Hempel's DN style explanation[21] whereas Batterman's view of understanding challenges the DN style explanation as well as its claim to scientific understanding. The idea is this: the DN style approach can answer a question relating to the expectedness of a given phenomenon. For instance, a computer can demonstrate the expectedness of an explanandum which can be done by reading the results of computation. However just reading results and calculating the expectedness, may not constitute a proper understanding. We might want to know why, in general, patterns of certain types are expected. To cast the question in familiar terms: "why" does this universal behaviour occur? The "why" question consists of the three questions outlined in the first section and is made clearer in my discussion of Batterman's minimal model understanding versus Kitcher's unification account below.

[21] Though it should be noted that emerging from the D-N view is not necessarily true of some contemporary views.

My discussion proceeds thus: first, I will briefly characterise Kitcher's unification, then I will outline the disagreements between Batterman and Kitcher and then show how these disagreements can be resolved in part by a fuller characterisation of my version of Batterman's minimal model understanding.

I will pick out Kitcher's version of unificationist understanding and contrast it with Batterman's account of understanding using asymptotic analyses, because the global unificationist understanding from Kitcher comes closer to adopting the attitude expressed by the Batterman sentiment than the causal mechanical approach that Wesley Salmon refers to --- that is, to understand what is going on, to 'elucidate the crucial features of the problem'. Moreover, just by solving, for instance, Schrodinger's equation and the answers that come out of it in a certain way, we still would not be closer to understanding "why" they came out in that way. In this way, unificationist understanding, by reducing the number and types of argument, at least gives us understanding that the world is organised in a certain way, thereby reducing the number of types of facts we must accept as brute. Wesley Salmon explains unification as top-down understanding making as small a number as possible of independent assumptions to explain the way things are in the world. The explanations in Kitcher's unificationist conception of understanding are meant to organise knowledge in the most coherent and efficient way possible. In this way we have a world picture, a top-down picture, and we see how our experiences as well as different aspects of the world fit into this world picture.

A unification theory can do duty for scientific understanding insofar that its aim is understanding--by reducing the number of independent assumptions one must make when explaining the way things are in the world. This concept of unification is a global concept because understanding is achieved in possessing a world picture and seeing how different aspects of the world as well as our experience fit into this picture. Philip Kitcher's (1989) version of unification can be summarised as an attempt to derive as many as possible different conclusions using as small a set as possible of

patterns of derivation. Understanding is gained in terms of showing how to derive descriptions of disparate phenomena whilst using the same patterns of derivation, and given this, thus making us accept a smaller number of brute facts.

Batterman's account of understanding is clearly similar to, but also different from, that which Kitcher emphasises. The main difference is that Batterman's asymptotic analyses have a local dimension; paradigmatic unification theories in contrast, are global in scope. Kitcher's unification *starts* with a class of things or a set of axioms and show that particular things belong in that class, thereby unifying them; for example, an explanation for a certain phenomenon on this picture would belong in the class of electromagnetic phenomenon because you can derive the phenomenon from the same set of electromagnetic equations or axioms. In contrast to this picture, Batterman with his asymptotic analyses turns Kitcher's picture upside down to unify disparate phenomena: we unify things, encoding them in a shared universality class. So, for Batterman, this phenomenon belongs in the universality class for a reason, which is demonstrated by the stability of related mathematical structures and appropriate mathematical spaces e.g. by renormalisation group methods. Unifying things in a certain universality class in this way is different to the way Kitcher shows; that which unifies is identified through the analysis itself. In short there are different reasons as to why one would unify things in different classes for each author.

The unification view does not respect the particularity or individuality of the problem. Yet the understanding gained from asymptotic analyses is local in this way: given a problem like the shape of droplets breaking, e.g. dripping from a faucet, one has an explanation given by the mathematical emergent asymptotic structures e.g. the similarity solutions to the one-dimensional Navier Stokes equations describing the universality of the shape of breaking drops. The structures are differentiated given that the problem class forms a natural group -- for example, the problem class forms a natural [universal] group of breaking drops. Each problem in the problem class needs

detailed local consideration of the relevant physics e.g. the physics of the interface involved in water dripping from a faucet. So, this is how Batterman's account is differentiated from Kitcher's account.

To go deeper, even though a unificationist theorist can, in general terms, appeal to asymptotic analyses, the real explanation for asymptotic analyses comes from the mathematical display of the dominant features, like the existence of a similarity solution of the one-dimensional Navier Stokes equations for free surface flows---the similarity solution describes the shape function of drops dripping from a faucet as a member of the universality class for breaking drops , or in the case of the sex ratio model, the dominant feature is the substitution cost and the mathematical display of the dominant emergent feature is found in delimiting the sex ratio for natural populations as a universality class. As Batterman says "the explanations are really provided by the mathematically transparent display of the dominant emergent features…And these features may differ dramatically from case to case" (Batterman 2000 p.253). The mathematically transparent display of dominant features varies with respect to each case for instance, in the case of delimiting a universality class for the minimal model of breaking drops and the minimal model of Fisher's sex ratio model. It is local because the modeller uses his intuition and expertise *peculiar to the case at hand* regarding the particular mathematical display of the dominant feature. This is the real difference between unification understanding and the understanding using asymptotic analyses. As Batterman claims, his asymptotic analysis "is quite distinct from the unificationist theorists' very global aim of reducing using the number of argument patterns required to "unify" one scientific knowledge" (Batterman 2000 p.252).

To illustrate the difference, I will now refer to the example of breaking drops. A Kitcher unificationist explanation of water drop shape would use argument patterns of micro-physics that are used to explain many other things since he wants to organise the drop shape under a small set of laws. Here, the explanation unifies the universal water droplet shape with other things by using the same argument pattern to explain it as to explain the other things. The explanation does not itself

reduce the number of argument patterns, rather it is explanatory because it uses the same patterns used for other things. It should be noted that Kitcher and Batterman both would want to give a micro-physical explanation. But Batterman explains by demonstrating mathematically emergent structures that display the dominant features of the phenomenon's behaviour, like the emergence of the substitution cost as a dominant feature in Fisher's sex ratio model, as a minimal model. So, in this way, though Batterman's explanation involves micro-physical explanation, the explanandum is, in part, constituted by macroscopic features as we saw in the characterisation I gave earlier. The difference in the two types of explanation has important consequences for the difference in the different theories of understanding that each author would subscribe to.

It should also be noted that Kitcher's unification is symmetric and Batterman's is asymmetric. The reason why a theory won't come along and unify a la Kitcher and replace it is because the derivations using the models involve local assumptions and can't be derived from by conditions from anything more fundamental, and that is why we have consistency proofs. There may be a future where we can derive this from more fundamental principles.

My view is that the differences between Kitcher's unificationist account of understanding and Batterman's minimal model understanding can be reconciled in the following way: minimal model understanding can be further characterised as a *global concept of understanding localised*. Batterman's conception of understanding is unificationist or global, insofar as asymptotic analyses reduce the number of types of argument used to organise or unify disparate phenomena (because they form a natural problem class- see above) ---- like water droplets dripping from a faucet or the water droplets forming from a crashing wave and characterising them as belonging to the same universality class. However asymptotic analyses used in minimal model understanding are also local in the following way: each particular problem, like the shape of water dripping from a faucet, in the problem class e.g. droplets in general, requires a unique consideration of the particular relevant physics. In this sense, the consideration that culminates in the mathematical derivation is a local

affair. And a Batterman explanation is not organising the drop shape merely under a small set of laws as Kitcher requires, rather, it is also demonstrating emergent structures. This latter is a local affair because each problem demands specific consideration and intuition, for example with respect to the physics involved and the physicist's intuition that are used in the example of breaking drops.

There are two central features to note coming from the above discussion:

1. Batterman's asymptotic analyses provide unificationist understanding because asymptotic analyses reduce the number of types of argument used to organise or unify disparate phenomena and put them in the same universality class.

2. Asymptotic analyses are also local insofar as each problem requires specific consideration of the physics involved. In particular the asymptotic structures will vary, though the problem is naturally in the same class.

We can see both of these in the example of criticality (above). 1. is satisfied by considering a curve on a graph that plots temperature versus density from eight different fluids in reduced dimensional co-ordinates. The curve on the graph provides the different densities of liquid and vapour phases at different temperatures. The shape of the representative curve is the same for each fluid (each with different molecular composition). Therefore, the curve unifies these disparate phenomena and puts them in the same universality class. 2. is satisfied because the different fluids need to be represented in reduced dimensional co-ordinates and this requires careful consideration of the specific physics involved. Therefore the mathematical derivation involved in asymptotic analyses is a local affair: there are principled reasons grounded in the fundamental physics of the system as to

why many of the details that distinguish systems from one another are irrelevant when it comes to explaining the universal behaviour of interest.

So, a complete characterisation of minimal model understanding using asymptotic analyses requires one to consider both the bottom-up approach, the local aspect, as well as a top-down approach, the global aspect. Therefore, asymptotic analyses can be said to provide a species of understanding as unification, but where unification is achieved differently than in Kitcher's account.

5.4. Conclusions

In this chapter I articulated a general version of Batterman's method of minimal model explanation using paradigmatic renormalisation group methods. Then I presented a more detailed example from physics as well as an example from biology to show that Batterman's minimal model explanation can be applicable to other fields in science. I also gave Batterman's reasons for minimal model explanation being better than mechanistic. Lastly, I linked minimal model explanation to understanding discussing unification and contrasting this with the mechanistic variety. As a result, I have characterised Batterman's minimal model understanding as subscribing to a kind of "local" unificationist understanding.

6. Simple models facilitate understanding--the case for target-less models.

6.1 Introduction

In this chapter I will shift to a different sort of simple model and a different sort of understanding. I will examine how *target-less models facilitate objectual understanding, where the immediate object of understanding is a theory.* My discussion will draw from Luczak's (2017) analysis of the Kac Ring model. This case will be used to illustrate one way in which target-less models, via the agents using the model and by the elucidation of concepts, can facilitate the objectual understanding of a theory – in this case, statistical mechanics. I am interested in the model providing objectual understanding. I will do this by setting out four conditions that are necessary for what I have called 'objectual understanding' and by showing how the Kac ring model, or the agent using the Kac ring model, satisfies these conditions.

To recap, from chapter 2,

- a *target-less* model is a model that does not have a representational target.

It follows that a target-less model is also not an idealised representation of any such target. The model is not taken to serve a representational function because the scientists using it do not intend

it to represent a specific system or phenomenon. Instead of a representational relation, oftentimes the relation of similarity between theory and the model giving objectual understanding is exploited to do a different job. This is the case for the Kac ring, which can be used both (a) to draw inferences about the theory of statistical mechanics and (b) to point up similarities between the Kac ring and Boltzmann's gas, which are considered similar in virtue of instantiating some of the same properties. However, I will concentrate on the former (a) drawing inferences about the theory of statistical mechanics) because I am interested, in this chapter, in objectual understanding of the theory of statistical mechanics and not understanding a phenomenon such as Boltzmann's gas.[22]

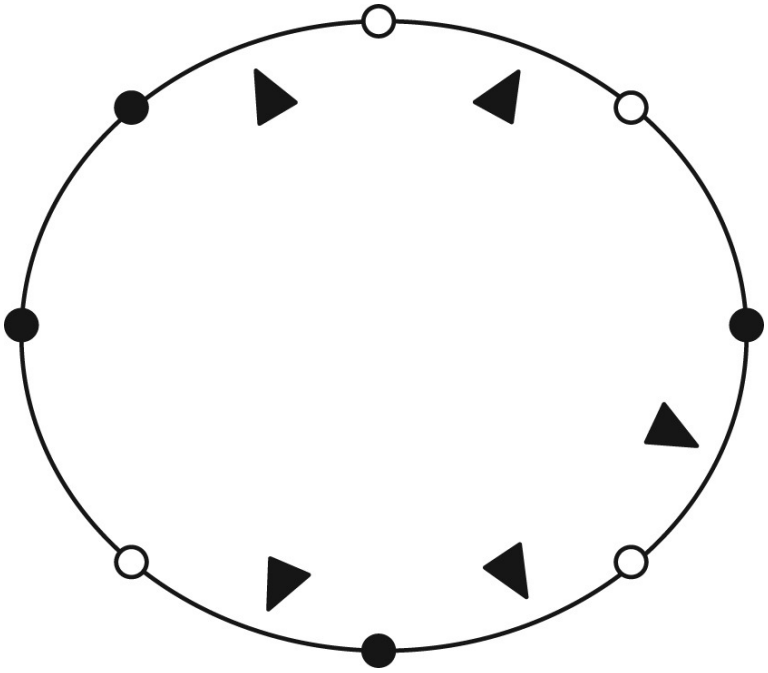
In section 6.2., I will introduce the Kac ring. Section 6.3. reintroduces objectual understanding (following on from chapter 3). Section 6.4. discusses one of the functions of the Kac ring: that, via the recurrence and reversibility objections to Boltzmann's theory of statistical mechanics, to gain a better understanding of statistical mechanical theory itself. After elucidating this function of the Kac ring, I will show how this function relates to objectual understanding in terms of satisfying four conditions.

6.2. The Kac ring and the recurrence and reversibility objections

The Kac ring model itself is simple and solvable. The model depicts N sites that are arranged in a circle. Sites are joined to their neighbour by an edge and some of the edges have a marker. Each site has either a white ball or a black ball.

[22] Understanding a phenomenon usually involves explanatory understanding which has been elaborated upon in the previous two chapters.

Figure 6.1. A Kac ring with N=8 lattice sites and n=5 markers (after Luczak 2017 figure 1)



The balls and markers are analogous to the molecules that make up a dilute gas. The evolution of the model proceeds in the following way: it involves a discrete set of ticks from t to t+1 by each ball moving in a clockwise direction to its neighbour; when a ball passes a marker, it changes colour analogous to the change in velocity of a molecule colliding with another molecule.

It helps to introduce some equations to show the model's dynamics. Let B(t) represent the total number of black balls and b(t) the number of black balls that pass a marker on the tick at t. And, let W(t) represent the number of white balls and w(t) the number of white balls that pass a marker on the tick at t. It follows that

(1) $B(t+1) = B(t) + w(t) - b(t)$

And

(2) $W(t+1) = W(t) + b(t) - w(t)$.

Also, one can study the difference between the number of black and white balls at different times:

(3) $\Delta(t)=W(t)-B(t)$ etc.

(Luczak 2017 pp.3-4)

Features or properties that comprise the Kac ring are:

- Its micro- dynamics are symmetrical under time reversal. Any reversed sequence of states is compatible with the system's micro-dynamics.

- The system is strictly periodic and so displays recurrence (Luczak 2017 p.3).

Recurrence is the feature that after a series of ticks, the ring returns to its initial state: i.e., there is some N such that -after N ticks, each ball reaches its initial state, having changed colour n [a number of] times. Moreover, if n is even, the number of ticks for recurrence when the colour changes n times is even; then if n is odd, it recurs after 2N ticks at most.

The Kac ring was first used by Mark Kac in 1959 in a series of lectures to introduce statistical mechanics and its applications. Specifically, in one lecture, he used the Kac ring model to introduce the statistical mechanical treatment of irreversible phenomena. This is the subject of this chapter, to show how the Kac ring model facilitates the objectual understanding of concepts in statistical mechanics.

Boltzmann's asymmetric result

I will firstly follow Luczak 2017 who, in the first instance, articulates the objections to the Boltzmann's story which the Kac ring elucidates. Boltzmann elucidated ideas relevant to statistical mechanics and thermodynamics, however the reversibility and recurrence objections, both important statistical mechanical concepts, criticised Boltzmann's ideas. I will elaborate both on Boltzmann's asymmetric result and the objections that follow in the statistical mechanical context.

Luczak stipulates that a thermodynamic system is typically an isolated macroscopic system which behaves thermodynamically if it is in equilibrium or spontaneously approaching equilibrium. The aim of statistical mechanics is to account for this behaviour with reference to the underlying micro-constituents. Such statistical mechanical systems typically have a micro-dynamic that is symmetric under time reversal and display recurrence if they have a bounded phase space energy hyper-surface, i.e., those systems that have a fixed energy. So, if one makes use of Poincare's recurrence theorem, then it is the case that if one considers a small open neighbourhood of a system's initial state, it will return to that state after leaving it for almost all initial phase space-points (paraphrased from Luczak 2017 p.3).

Boltzmann's asymmetric result can be derived from an attempt to account for thermodynamic behaviour in a classical framework, i.e. using Newton's laws of motion. Boltzmann argued for a unique distribution of the velocities of molecules in a contained dilute gas that changed under collisions. This was the Maxwell-Boltzmann distribution which was stable under collisions. Furthermore, he claimed that gases with different initial distributions all moved towards the Maxwell-Boltzmann distribution. To this end, his argument posited a quantity H that reached the minimum value for the distribution and argued that it monotonically decreased to its minimum, giving us Boltzmann's H theorem. Boltzmann's H theorem is a consequence of his transport equations which track the system's behaviour, specifically its approach to equilibrium. Boltzmann's H theorem is a temporally asymmetric result (paraphrased from Luczak 2017 p.3).

Criticism of this result queried how Boltzmann got his temporally asymmetric H theorem from the system's micro-dynamics that was assumed as symmetrical under time reversal, etc. A couple of objections showed that he could not have derived the H theorem with the assumption that the system's dynamics was symmetric under time reversal (paraphrased from Luczak 2017 p.3).

These objections are the reversibility and recurrence objections. The reversibility objection applies to those systems that have micro-dynamics that are symmetrical under time reversal. This is the case for any set of trajectories of molecules of a dilutes gas: that the time reversal of that set of trajectories is compatible with the gas's or system's dynamics. The second objection, the recurrence objection, also applies to Boltzmann's gas (a dilute gas) insofar as it applies to classical systems with bounded phase space energy hyper-surfaces. The objection goes that if we consider a small open neighbourhood of the system's initial state and then ask if that system, during its evolution, will return to that state after leaving it, then the answer by Poincare's recurrence theorem is positive for all initial phase points (apart from a set of Lesbegue measure zero). This implies that there is no monotonic decrease, given that the system will return to its initial state, of Boltzmann's quantity H (closely paraphrased from Luczak 2017 p.3).

Therefore, the objections culminate in saying that not all initial micro-states of the gas at any time lead to a monotonic decrease of Boltzmann's H quantity. These objections shows a straightforward contradiction to the basic features of Boltzmann's gas given his H theorem. And so, the H theorem is in contradiction to the consequences of the system's micro-dynamics---that system that Boltzmann was interested in i.e. a dilute gas. Therefore, his approach to equilibrium could not have been ascertained given the system's micro-dynamics (paraphrased from Luczak 2017 p.3).

The story goes on, with another question: how then did Boltzmann derive his asymmetric result? The answer is that we need more than the application of Newton's laws of motion to molecular collisions to get this result. In Boltzmann's case, it was with the addition of the Stoßzahlansatz assumption that gave the temporally asymmetric result (Boltzmann's H theorem). The

StoBzahlansatz is an asymmetric assumption that is applied in the derivation of Boltzmann's transport equations (tracking the system's behaviour). The assumption says that there is always an absence of correlations between the velocities of colliding molecules. So, it is not the case that for all times the system's micro-dynamics are symmetrical under time reversal and that for all phase space points, Poincare's recurrence theorem holds. In this way we can see how Boltzmann could have derived his H theorem. But as we know, the objections denied Boltzmann's derivation of his H-theorem given the system's micro-dynamics.

The application of the model of the Kac ring

In the context of the Kac ring, we say that a system is in equilibrium when the total number of black balls are equal to the total number of white balls, (W(t)=B(t)). W and B and Δ in equation (3) are macroscopic quantities though there are many different microstates that give rise to the same macroscopic quantity. However, w and b (see equations (2) and (3)) are localised bits of information about individual sites and as such cannot be calculated without knowing the location of each marker and the colour of each ball at each site. Moreover, the evolution of the macroscopic quantities cannot solely be determined by macroscopic state information alone.

This limitation is overcome if we make an analogue of the above non-dynamical assumption (the Stoßzahlansatz assumption): we suppose that the fraction of the white or black balls that change colour in each tick is equal to the probability that an edge has a marker where the probability is equal to the number of markers divided by the number of edges. This is a non-dynamical assumption and is analogous to assuming that the Stoßzahlansatz holds at all times. So, the colour of each ball, at each tick, is probabilistically independent of whether there is a marker in front of it. This

indicates an absence of correlation and introduces a temporally asymmetric element so that some sequences of states that are compatible with the analogous assumption are not compatible with their time reverse.

To give an example using the Kac ring, showing the success of the analogue of Boltzmann's H theorem with an analogue of the Stoßzahlansatz assumption: suppose we consider a ring with only white balls at each site and a random distribution of markers. We let the system evolve for one tick and the balls that pass the marker change colour, then the analogue assumption which holds for this sequence of states does not hold for its time reverse because in this case, the colour of the balls are not correlated with location of markers. Black balls are only found at sites that have a marker in front of them. With the analogue assumption we can yield a transport equation for the system that tracks the system's behaviour including the approach to equilibrium, which is analogous to Boltzmann's H result (paraphrased from Luczak 2017 p.4).

However, this result is inconsistent with the system's micro-dynamics which are symmetrical under time reversal, so it is not the case that all micro-states of the system are compatible with their macroscopic properties at any time leading to a monotonic decrease of the quantity appearing in the transport equation. This is an instance of the reversibility objection. The system also instantiates the features or properties needed for the recurrence objection because it is strictly periodic so that no initial microstate yields a monotonic decrease of the quantity appearing in the transport equation. As will be discussed further below, the above application of the ring model is to understand more clearly the objections and their implications in relation to other ideas.

Informally, we can describe the situation as follows: the Kac ring is used to instantiate the objections because it is a simple model that has the requisite properties[23], and with the added analogue of the Stoßzahlansatz assumption, we can understand more clearly the nature and significance of the recurrence and reversibility objections to statistical mechanical theory. We can

[23] The properties of the Kac ring include its dynamics being symmetrical under time reversal.

see how the objections arise. Moreover, although the objections were originally pitched against Boltzmann's dilute gas, they are not specific to it; the ring's analogue objections can be pitched against any system that instantiates the required properties. In what follows, I will develop a more detailed account of the understanding that is gained, characterising it as a kind of objectual understanding.

6.3. Objectual understanding

- The *object of objectual understanding*- that which is to be understood- can be a body of knowledge or bodies of information that may be a theory or part of a theory or even theoretical concepts; it can also be a complex (physical) system e.g., the climate system (see the example below).

In general, objectual understanding as I discuss it is given by a model. The model is the vehicle by which one can understand the object of objectual understanding, so objectual understanding involves:

1. The vehicle of objectual understanding, that in my case is the model.
2. Providing a kind of understanding and
3. The thing that is to be understood.

One function of the Kac ring model is to supply objectual understanding of statistical mechanics. Recall that the Kac ring model is not representational, it has a relation of similarity to the theory of statistical mechanics.

One can follow Elgin in explicating the idea that objectual understanding is only approximate. She requires that the vehicle of understanding --- in my case, a model--- be 'tethered' to the facts. A tether is a relation between a model and the body of information, which is the object of objectual understanding that is weaker than 'being true of'. In the case of false or idealised[24] models that supply objectual understanding, the idealizations in it are felicitous falsehoods. They are felicitous just because they get epistemic access to the facts that would be otherwise hard to obtain. They do this in virtue of exemplifying certain features shared with the facts. An exemplar that exemplifies a certain property is one that both refers to it and instantiates it; for further elaboration see chapter 3. Elgin also thinks that objectual understanding comes in degrees (Elgin 2007).

Kvanvig (2003) contra Elgin, says that we understand if most of the, propositions (especially most of the central ones) that make up a coherent take (the vehicle of understanding) on the object of objectual understanding are true, so understanding for Kvanvig is at least weakly factive, (see chapter 3). My account developed from Baumberger (2019) consisting in the 4 conditions below, is in opposition to Kvanvig---I follow Elgin and allow that models that are not representationally accurate can nevertheless provide objectual understanding, where, in the case studied here, the model is the Kac ring and it provides understanding of theory.

Another conception of objectual understanding, one that (like Elgin's) also comes in degrees is from Baumberger (2019). According to him, an epistemic agent A understands the subject matter S

[24] Luczak does not concern himself with idealised models and I can only assume that he and Elgin use different definitions of idealisation but the general message, common to both, is that false or target-less models which are neither true nor false models can give understanding.

(i.e., the object of understanding) by means of a theory or model T (the vehicle of understanding) only if A commits herself sufficiently to T of S and to a sufficient degree:

1. A grasps T
2. T answers to the facts and
3. A's commitment to T of S is justified. (Baumberger 2019 p.376)

It is context dependent as to how well the conditions are met for complete understanding i.e. for an attribution of understanding coming out as true. Baumberger goes on to stipulate that there is only a small possibility that there will be a context in which satisfaction of the conditions will lead to an ascription of outright understanding to the agent. His intuition is that grasping, rightness (answering to the facts) and justification are "good-making" features of understanding rather than prima facie necessary conditions. "Good-making" features of understanding are referred to as evaluative dimensions of understanding and the evaluative dimensions allow for objectual understanding to come in degrees.

It is important to stress that this is one characterisation of a feature of objectual understanding – that objectual understanding comes in degrees. For Baumberger (2019), determining how good an agent's understanding is depends on informal evaluation and judgement, (see Kuhn's (1977) theory choice); so, we cannot give a rule for saying what different degrees of understanding involve. This is because the conditions for attribution are open for interpretation and require a deeper explication. There is no one correct procedure for assessment concerning how well each of the conditions are met.

The content of objectual understanding can also be illustrated by the following example about climate change which is compatible with the above explication:

Scientist S claims to understand climate change by means of a climate model, where the equations of the model are derived from physics and with the help of empirical assumptions about factors such as cloud albedo. (Baumberger 2019 p.374, Baumberger and Brun 2016 p.3)

To defend the claim that a scientist understands climate change: it must be that the model answers to the facts or reflect the facts. This may include idealised models given certain qualifications. The scientist must also grasp the model by way of using it to answer a wide range of questions about climate change. Quality of comprehension of how the model behaviour comes about from its various components (e.g., physical principles, parameterisation's etc.) is needed for the scientist to be able to efficiently use the model for predictive and explanatory purposes. Finally, the scientist must provide reasons speaking in favour of the model and show that the model is good in accommodating evidence, precision of results, explanatory power etc. This then, satisfies the justification condition.

I elaborate on the above explication and example by articulating four conditions for a better characterisation of objectual understanding, drawing directly on the discussion in Baumberger 2019. I keep my characterisation as general as possible to allow for greater scope. I also follow Baumberger and Brun (2016) and stipulate that each of the four conditions listed below is necessary for objectual understanding. Note that this is different from Baumberger (2019).

1. The *commitment condition* is an attitude, of an agent, regarding some content, e.g. the model. This attitude according to Baumberger is epistemic acceptance. *Commitment* in terms of acceptance is to treat the model, call it P, with the attitude 'given that P', as

opposed to 'believing that P'. *Acceptance* is to adopt 'a policy that includes P as one of its premises for deciding what to do or think in a particular context, whether or not one feels it to be true that P', (Cohen 1992 quoted by Baumberger 2019 p.384). Epistemic acceptance is cashed out in terms of an agent taking the theory or model to be useful given specific epistemic purposes. Such purposes can include prediction, retrodiction and explanation. These purposes, further, are based on an evaluation of the theory's or model's performance regarding its epistemic goals. Commitment, as Baumberger (2019) sees it, is a necessary condition, and cannot be taken in a weaker sense i.e., as an evaluative condition for understanding for an account of understanding that comes in degrees.

2. *Answering to the facts in terms of similarity* is the second condition to be met for understanding. This means that the theory or model that supplies understanding answers to the facts that are true of the object of objectual understanding, to the degree that the object as depicted by the theory or model is in relevant respects similar to the way that the object really is. Which respects are relevant depends on the context and purpose that the theory or model is intended to serve. Most authors accept answering to the facts or a rightness condition for objectual understanding. If we stipulate that even the most radically false idealised models can give some understanding, the quality of understanding can ceteris paribus be a function of how well it answers to the facts. Baumberger requires answering to the facts, however Elgin tries to develop a looser requirement and it should be noted that where Elgin is defensible, one may want to substitute hers for Baumberger's stronger requirement as stated here.

3. The third condition is that the *agent grasps* the model. The condition of grasping is the ability to use a theory or model implying the ability to draw consequences of the theory or model concerning aspects of the subject matter/ body of information. On one account,

(Baumberger 2019), this need not involve deduction, but something called plausible reasoning. A good grasp of the model is the ability to solve qualitative problems using the model. To do this, one draws out the consequences of the model without exact calculations generating knowledge of the model behaviour that emerges from the interaction of the components of the model. This is done before, or instead of, being able to solve quantitative problems---- especially if it is the case that the quantitative problems are very complex, i.e. unsolvable analytically. Moreover, a good grasp of the theory/model may also involve the ability to assess conditions and limits of application to the object of understanding. It does not matter which theory of grasp you subscribe to. If some understanding is possible, and one grasps what is to be understood, the quality of your understanding ceteris paribus is a function of your grasping the theory/body of information or model or how well you grasp it.

4. The fourth is a *justification condition*. Standardly, accounts of epistemic justification of knowledge and belief require that the justifiers be truth conducive, but this may not be applicable for understanding, given the idealisation/simplicity of models (paraphrased from Baumberger and Brun 2016 p.4).

I adapt Baumberger (2019) by giving the following proposal: whether an agent's commitment to a theory is justified depends on (a) if the theory/model is internally consistent and externally coherent with background theories and assumptions, (b) if the theory/model accommodates available evidence e.g., intuitions in the case of non-empirical theories/models or observation-based data in the case of empirical theories/models. (c) if the theory/model furthers epistemic goals including e.g., simplicity, fruitfulness, explanatory power etc. (d) if the agent can assess how well (a) to (c) are met. And (e) is the reliability of the model evaluation or how reliable the evaluator of the model is.

> To elaborate, dimensions (a) and (b) assess how well the agent's theory or model answers to the facts about the object of understanding. Epistemic goals in (c) are used to assess the systematicity of the vehicle of understanding, the theory or model, its intelligibility for the agent, or the ease by which the agent uses the theory or model and relevance to a problem. (d) assesses how well (a) to (c) are met. (e) is how unlikely it is that the evaluation leads the agent to commit herself to a theory or model not answering to the facts or not being relevant to the problem at hand. So, dimensions (a) to (e), help explicate a justification condition for objectual understanding.

However, if we are to stick to a theory of objectual understanding that comes in degrees, all the conditions apart from the commitment condition can be construed as constituting only evaluative dimensions. Then, the extent of an agent's understanding depends on how well the agent does at fulfilling the conditions which thus determine the quality of understanding. However, even on this view, if the commitment condition comes in degrees, the extent of understanding is not dependent on the degree of commitment to the theory/model she is committed to---- Baumberger (2019) wants to stipulate that the commitment condition, at the very least, is necessary.[25] However, Baumberger and Brun (2016) stipulate that each condition may be necessary, i.e., they are at least met to some minimal extent, and beyond this, we simply get better understanding. I will adopt this latter view because it allows for greater breadth in terms of application-- the view that I am adopting still allows for objectual understanding to come in degrees.

6.4. The Kac ring and objectual understanding

[25] This is the case even for Baumberger's theory of objectual understanding in degrees.

In the case of the Kac ring, the kind of understanding supplied is objectual understanding in so far as the vehicle of understanding goes some way to providing an agent an understanding of a topic or subject matter. With the Kac ring, we are making progress towards objectual understanding of the theory of statistical mechanics; and if the objections obtain, that these objections also hold for Boltzmann's dilute gas in the sense that we learn something about the dilute gas. In effect the objections are objections to Boltzmann's reasoning. We come to this latter conclusion by reasoning from the similarity that obtains with Boltzmann's dilute gas also exhibiting the above properties (its micro-dynamics) and assumption. So, in this case we understand a phenomenon as well. As the main point I am making is that of objectual understanding, I shall confine myself mostly to the statistical mechanical theory as the object of understanding. We get objectual understanding by using the model to show how the objections arise. In the case of the climate example above, it can be adapted to show this insofar as we gain objectual understanding of the climate system by answering questions etc., about the system using the climate model.

The above adapted schema of four conditions (section 6.3) by Baumberger (2019) is fulfilled by an agent using the Kac ring to get objectual understanding via the two objections to the theory of statistical mechanics:

1. To re-cap, the commitment condition is the attitude of epistemic acceptance cashed out in terms of the agent being committed to the model. Epistemic acceptance is the extent to which the agent accepts the model as useful for certain epistemic purposes. These purposes are based on an evaluation of the model's performance regarding its epistemic goals.

2. To re-cap, Boltzmann's gas is the Kac ring's secondary object of understanding. Its primary object is the theory of statistical mechanics. We can say that the Kac ring goes all the way to answering to the facts because in this case, this means answering to the facts about statistical mechanical theory, in particular the facts about the statistical mechanical account of Boltzmann's gas.

This way of thinking about answering to the facts is analogous to a situation in which we wanted to come to understand something about, for instance, aether theory, using a simple model. We would need the model to answer to the facts about aether theory rather than real aether. The reality of the aether does not matter. In our case, the Kac ring answers to the facts because the ring instantiates the same set of properties that are found in a statistical mechanical system in the theory of statistical mechanics; those that along with the Stoßzahlansatz assumption, ground the objections.

In particular, the reversibility and recurrence objections are derived from the system's micro-dynamics and an added assumption. Analogues to these are found in the Kac ring. The reversibility objection applies to systems whose micro-dynamics are symmetric under time reversal, and the recurrence objection is grounded by a classical system with a bounded energy hyper-surface. Via the Kac ring analogue, through the relation of similarity between the model and a secondary object of understanding, these analogue objections also applies to the secondary object of understanding: real systems with a total fixed energy, for example Boltzmann's gas (paraphrased from Luczak 2017 p.6). We say that for both the primary and secondary objects of understanding, the Kac ring model answers to the facts, and this model contributes to objectual understanding of the recurrence and reversibility objections in statistical mechanics.

3. The grasping condition is fulfilled just in case the agent grasps the model. To re-cap, here, grasping it is cashed out in terms of making use of the model (Baumberger 2019 p. 379): she grasps the model to the extent she is able to use t the model i.e., the Kac ring, to derive the recurrence and reversibility objections in statistical mechanics. A good grasp of the model involves the ability to solve qualitative problems with the model if the model cannot be or is too complex to be solved analytically. However, this is not the case with the Kac ring; one can give quantitative solutions to the problem with ease, i.e. the solution that the analogue statistical mechanical objections hold in the ring. We can say we have a good grasp of the Kac ring model if we see how the model's behaviour emerges from its components. Also note that the above "use" notion of grasp is peculiar to Baumberger's (2019) conception of grasp for objectual understanding. But to reiterate, grasping cashed out in terms of use or abilities is not necessary for the condition of grasping to be applicable i.e. it does not matter which theory of grasp you subscribe to; the condition of grasping of some sort for an objectual theory of understanding remains. Indeed, in chapter 3 I have elaborated on other theories of grasp that are more dominant in the literature. Following the general intuition in chapter 3 we can say that grasping the Kac ring model is a first step in achieving objectual understanding of the objections in statistical mechanics.

4. The justification condition is fulfilled just in case there are sufficient reasons that speak in favour of the model's use. This roughly means that an agent's commitment is based on reasons which speak in favour of the model (Baumberger 2019 p. 384). Dimensions (a) and (b) (from 6.3 above) assess how well the model answers to the facts, and from the discussion above (number 2 of these applications of the four conditions), I conclude that the model answers to the facts, given that the ring model instantiates the same

properties and has the same analogue assumption as in the theory of statistical mechanics. Dimension (c) assesses the systematicity of the model in terms of furthering the agent's epistemic goals and how easy it is for the agent to use the model. The level of systematicity is high with the Kac ring, it is very simple and highly fruitful, making it easy to see the objections in statistical mechanical theory thus furthering the agent's epistemic goals. Furthermore dimension (d) is met because the agent can indeed assess how well dimensions (a) – (c) are met. Dimension (e) assesses the unlikeliness of the evaluation not answering to the facts or not being relevant to the problem at hand. We can say that the evaluation process does indeed let the agent commit herself to a model that sufficiently answers to the facts of the objections to statistical mechanical theory. Therefore, we can say that the Kac ring fulfils the justification condition that is dimensions (a) – (e) above.

The Kac ring in effect helps the agent see the need to assume something extra in the theory of statistical mechanics in order for Boltzmann's results to go through. What she comes to understand about the theory is that we need additional assumptions for the behaviour that we are interested in to be successfully derived, therefore appreciating the role of those assumptions needed.

The objections engage with two foundational statistical mechanical concepts- recurrence and reversibility; they are important historically, but they may be difficult to grasp or be fully appreciated by students that are unfamiliar to statistical mechanics. With the help of the Kac ring, many students can grasp and appreciate the objections and their implications and how they relate to other ideas in statistical mechanical theory, advancing their objectual understanding of the theory (Luczak 2017).

6.5. Conclusions

In this chapter, I have shown one way in which the Kac ring can contribute to objectual understanding of the theory of statistical mechanics. I have shown that the Kac ring, as a target-less simple model, elucidates the recurrence and reversibility objections in statistical mechanics, and in doing this fulfils the above four conditions (generalise) that I offer for objectual understanding, building from Baumberger (2019) and Baumberger and Brun (2016). The Kac ring model helps agents to get a better, objectual understanding via the objections (recurrence and reversibility objections), of the theory of statistical mechanics.

7. Conclusions

In this concluding chapter, I will outline the main contributions of each chapter in this thesis to the main project: how simple models can facilitate scientific understanding. Finally, I proceed to articulate the various ways each chapter can be extended in terms of further work.

7.1. Key points

The second chapter characterised 'simple models'. I identified four kinds of simple models: toy models, epistemic surrogates, minimal models, and target-less models. In doing so, I departed from the existing views on categorisation of such models insofar as I gave Batterman's minimal models their own category, distinguishing it from other models such as epistemic surrogate models (see Reutlinger et al. 2018, Weisberg 2007, 2013). My categories are not mutually exclusive or exhaustive but do highlight some important characteristics of various simple models.

The third chapter is also scene setting, this time focusing on understanding. I identified various kinds of understanding: understanding that, how, why, with and objectual as well as different conditions for understanding relevant to the present project. I adopt the view that understanding is non-factive because I think there are some cases in which we gain genuine understanding from largely false or idealised assumptions, even if often the basis is at least weakly factive. Along with chapter two, this discussion provided a foundation for the primary analysis of the thesis in later chapters, which explores how different kinds of simple models facilitate scientific understanding.

The fourth chapter began the investigation into simple models and understanding in earnest. It examined how simple models, by providing how-possibly and how-actually explanations, can facilitate scientific understanding. I discussed different kinds of how-actually and how-possibly explanations and outlined what understanding with simple models would look like for such explanations. The discussion built upon Reutlinger et al.'s (2018) analysis of understanding with toy models, appealing to their "refined simple view" of understanding. The conclusion of the chapter highlighted the challenge of identifying a dividing line between the how-actually and how-possibly explanations.

Chapter five focused on understanding via minimal model explanation as articulated by Batterman (2002). The key contribution of this chapter was to further articulate the character of understanding one can obtain from minimal model explanation. I proposed that Batterman's (2002) partial account of such understanding could be further enriched by characterising it as a kind of localised unification, where unification is achieved differently from the Kitcher's (1989) classic account.

Chapter 6 focused on target-less models as a kind of simple model to facilitate understanding. The distinctive feature of these kinds of models is that they have no representational targets. Following Luczak (2017), I used the example of the Kac ring model to show how target-less models facilitate objectual understanding, where the immediate object of understanding is a theory; the theory in question in this case was the theory of statistical mechanics via the objections to Boltzmann's theory of statistical mechanics. The main contribution of this chapter was in articulating four conditions for a characterisation of objectual understanding and then showing how those were met in the Kac ring case. Previous chapters focused on understanding "why", and this chapter aimed to show that simple models can facilitate other types of understanding as well.

7.2. Further work

Each of the aforementioned chapters presents opportunity for further work. Moving beyond chapter 2, we might investigate further how best to organise or categorise simple models and whether my definition of a simple model as very simple and highly idealised can be improved upon. The notion of idealisation as I defined it was broadly construed into two types, and this may be contentious. For instance, there might not be a coherent division between Aristotelian and Galilean idealisation, as noted by some scholars (Weisberg 2013).

Further work on chapter 3 would delve more deeply into the consequences of embracing non-factivism about understanding. There is also the question of whether my take on understanding which highlights grasping, non-factivity and explanation, can be fleshed out further in the service of analysing how simple models facilitate understanding.

Further work emerging in the light of chapter 4 could focus on investigation to the way in which how-possibly explanations can lower confidence in an impossibility hypothesis (Grune-Yanoff 2009). Some scholars such as Fumagalli (2016) respond to Grune-Yanoff (2009) that one cannot learn from minimal (my simple) models. Learning here is an interaction between the how-possibly explanation and the impossibility hypothesis. The idea that we gain scientific understanding from minimal models is thereby challenged if we cannot learn from these models. How to respond to objections like that of Fumagalli is a topic for future work.

Further work building on chapter 5 can include fleshing out my account of localised unification as minimal model understanding in greater depth and considering the consequences of adopting this stance. Further work could also include questioning the role of idealisation in Batterman's minimal

model explanations. Norton's (2012) critique says that there is nothing unique about minimal model explanations and that really the infinite idealisations that Batterman uses in his minimal model explanations are not the only way to understand the phenomena. In fact, adopting his view shows that there are no infinite idealisations used in these cases (Strevens 2019). Strevens then asks if there is a difference between the rules that govern idealisations in the non-infinite (simply idealised) cases and Batterman's asymptotic idealisations involving infinities. He shows that not all idealisations are simple idealisations and thus he preserves the Batterman intuition in the face of Norton's critique. Further work can include whether the two different kinds of idealisations -- simple and infinite--- can be compared with each other in the first instance.

Finally, with regard to the discussion of chapter 6, it would be interesting to consider whether there are other examples of simple target-less models that provide understanding of some sort. With the example of the Kac ring itself, it is also worth investigating if the Kac ring really is representationally target-less. Originally the Kac ring had Boltzmann's weak gas as an intended target; if so, then can we not say that one of the functions of the Kac ring is to represent Boltzmann's gas? In the example of the Kac ring model as elaborated by Luczak (2017), the representation relation is replaced by similarity relation which is established by an argument from analogy. We can then ask if similarity might be the only relation that could work in these cases; perhaps there could be other types of relations that can also work?

There are thus many exciting avenues for further work, but in this project, I have hoped to at least to have clarified, and bought together in a single discussion, a range of views on how simple models can facilitate scientific understanding.

Bibliography

- Batterman, R. (2009), 'Idealisation and modelling', *Synthese*, 169, 427-446.
- ----------(2002), *The devil in the details: asymptotic reasoning and explanation, reduction, and emergence,* Oxford University Press.
- ----------(2000), 'A modern (=Victorian) attitude to scientific understanding', *The Monist*, 83, 228-257.
- ----------(2018), 'Autonomy of theories: an explanatory problem', *Nous*, 52, 4, 858-873.
- Batterman, R. and Rice, C. (2014), 'Minimal model explanations', *Philosophy of science*, 81, 349-376.
- Baumberger, C. (2019), 'Explicating objectual understanding: Taking degrees seriously', *Journal for general philosophy of science*, 50, 3, 367-388.
- Baumberger, C. and Brun, G. (2016), 'Dimensions of objectual understanding' in Grimm, S., Baumberger, C. and Ammon, S. (eds) *Explaining understanding. New perspectives from epistemology and philosophy of science,* Taylor and Francis.
- Bokulich, A. (2014). 'How the Tiger Bush got its stripes: how possibly vs how actually model explanations', *The Monist*, 97, 3, 321-338.
- De Regt, H. (2017), '*Understanding scientific understanding'*, Oxford University Press.
- Elgin, C. (2007), Understanding the facts, *Philosophical studies* 132, 33-42.
- ----------(2017), 'Understanding', *Routledge encyclopaedia for philosophy*, Taylor and Francis.
- ----------(2004), 'True enough', *Philosophical issues*, 14, 113-131.
- ----------(2012), 'Understanding Tethers', in *Epistemology: contexts, values and disagreement*, Jager et al. (eds), Ontos Verlag.
- Elliott-Graves, A. and Weisberg, M. (2014), 'Idealisation' *Philosophy compass*, 9, 3, 176-185.
- Frigg, R.P. (2009), 'Models in Physics', *Routledge encyclopaedia of philosophy*, Taylor and Francis.
- Frigg, R.P. et al. (2015), 'Philosophy of climate science part 2- modelling climate change', *Philosophy compass*, 10, 12, 965-977.
- Frigg, R.P. and Hartmann, S. (2012), 'Models in science'*, Stanford encyclopaedia for philosophy*, The Metaphysics Research Lab, Centre for the Study of Language and Information, Stanford University.

- Fumagalli, R. (2016), 'Why we cannot learn from minimal models*, Erkenntnis*, 81, 433-455.
- Grune- Yanoff, T. (2009), 'Learning from minimal economic models*, Erkenntnis*, 70, 1, 81-99.
- Khalifa, K. (2019), 'Non-factive understanding: a statement and defence', *Journal for general philosophy of science,* 50, 3, 345-365.
- --------(2013), 'The role of explanation in understanding', *The British Journal for the philosophy of science,* 64, 1, 161-187.
- --------(2013), 'Is understanding explanatory or objectual*, Synthese* 190, 6, 1153-1171.
- Kitcher, P. (1989), 'Explanatory unification and the causal structure of the world' in '*Explanation and the causal structure of the world'* in Salmon, W. (ed), Princeton University Press.
- Kuhn, T.S. (1977), 'Objectivity, value judgement and theory choice' in '*The essential tension: selected studies in scientific tradition and change*', University of Chicago Press.
- Laymon, R. (1998), Idealizations, *Routledge encyclopaedia of philosophy*, Taylor and Francis.
- Lipton, P. (2009), 'Understanding without explanation', in *Scientific Understanding-philosophical perspectives,* Leonelli, S. et al. (eds), University of Pittsburgh Press.
- Luczak, J. (2017), 'Talk about toy models', *Studies in history and philosophy of modern physics,* 57, 1-8.
- McMullin, E. (1985), 'Galilean idealisation'*, Studies in history of philosophy of science,* 16, 247-243.
- Mizrahi, M. (2012), 'Idealisation's and scientific understanding', *Philosophical Studies,* 160, 2, 237-252.
- Norton, J. (2012), 'Approximation and idealization: why the difference matters', *Philosophy of science*, 79, 2, 207-232.
- Parker, W. (2014), 'Simulation and understanding in the study of weather and climate', *Perspectives on science*, 22, 3, 336-356.
- Potochnik, A. (2020), Idealisations and many aims, *Philosophy of science: proceedings of PSA,* 87, 5, 993-943.
- Reutlinger, A. et al. (2018) 'Understanding with toy models', *British Journal of Philosophy of Science,* 69, 1069-1099.
- Rice, C. (2019), 'Models don't decompose that way: A holistic view of idealised models', British *journal for philosophy of science,* 70, 179-208.
- ------(2018), 'Idealised models, holistic distortions, and universality'*, Synthese,* 208, 195, 2795-2819.

- -------(2016), ‘Factive scientific understanding without accurate representation’, *Biological philosophy,* 31, 81-102.
- Strevens, M. (2013), ‘No understanding without explanation’, *Studies in history and philosophy of science*, 44, 510-515.
- ---------(2020), ‘Grasp’, draft version, Michael Strevens, New York University.
- Sullivan, E. and Khalifa, K. (2019), ‘Idealisations and understanding: much ado about nothing’, *Australasian Journal of philosophy*, 94, 4, 673-689.
- Verrault–Julien, P. (2019), ‘How could models possibly provide how possibly explanations?’, *Studies in history and philosophy of science*, *part A,* 73, 22-33.
- Weisberg, M. (2007), ‘Three kinds of idealisation’, *The Journal of philosophy*, 104, 12, 639-659.
- ----------(2013), ‘Simulation and similarity’, Oxford University Press.
- Woody, A. (2013), ‘How the gas law is explanatory’, *Science and education*, 22, 7, 1563-1580.